

Л. В. Тарасов

МИР, построенный на вероятности



ratione salva manente; nemo enim convenientius allegabit
quam ut magnitudine orbium multitudo temporum metiatur; ordo sphae-
rarum sequitur in hunc modum a summo capientes initium.
Prima et suprema omnium est stellarum fixarum sphaera seipsam et omnia continens, ideoque immobilis, nempe universi locus, ad quem motus et positio caeterorum omnium syderum conferatur.

1 Stellarum fixarum sphaera immobilis
2 Saturnus xxx anno revolvitur
3 Jovis xii annorum revolutio
4 Martis bima revolutio
5 Telluris cum Luna an re
6 Veneris nonimestris
7 Merc lxxx dierum
Sol

Nam quod aliquo modo illam etiam mutari existimant aliqui: nos aliam, cur ita appareat, in deductione motus terrestris assignabimus causam. Sequitur errantium primus Saturnus: qui xxx anno suum complet circuitum. Post hunc Jupiter duodecennali revolutione mobilis. Deinde Mars, qui biennio circuit. Quarto in ordine annua revolutio locum obtinet: in quo terram cum orbe lunari tanquam epicyclo contineri diximus. Quinto loco Venus nono mense reducitur


m'
si
m
S1

Л. В. Тарасов

МИР, построенный на вероятности

Книга для учащихся

Москва
«Просвещение»
1984

ББК 20
Т19

Рецензенты: *академик АН УССР Б. В. ГНЕДЕНКО,*
д-р филос. наук Г. Я. МЯКИШЕВ и
кандидат физ.-мат. наук О. Ф. КАБАРДИН

Тарасов Л. В.
Т19 Мир, построенный на вероятности: Кн. для учащихся.— М.: Просвещение, 1984.—191 с., ил.

В книге в интересной и популярной форме рассказано о вероятностных причинно-следственных связях и их принципиальной роли в природе, научном познании мира, человеческой практике. Рассмотрены физические основы кибернетики и теории информации, молекулярно-кинетической теории и квантовой механики, генетики и эволюционной теории. Предназначена школьникам старших классов.

Т $\frac{4306021100-640}{103(03)-84}$ 211—84

ББК 20
5

...В природе, где как будто господствует случайность, мы давно уже установили в каждой отдельной области внутреннюю необходимость и закономерность, которые пробивают себе дорогу в рамках этой случайности.

Ф. Энгельс

Предисловие

Великое множество событий и явлений совершается в окружающем нас мире. События взаимосвязаны — одни из них являются следствием (исходом) других и, в свою очередь, служат причиной третьих. Вглядываясь в гигантский водоворот взаимосвязанных явлений, можно сделать два важных вывода. Во-первых, наряду с совершенно определенными, однозначными исходами встречаются неоднозначные исходы. Если первые можно предсказать точно, то вторые допускают лишь вероятностные предсказания. Другой не менее важный вывод состоит в том, что неоднозначные исходы встречаются значительно чаще, чем однозначные. Вы нажимаете кнопку, и стоящая на вашем столе лампа загорается. Здесь второе событие (загорелась лампа) является однозначным исходом первого события (нажата кнопка). Такое событие называют *строго детерминированным* (от латинского *determinare* — «определять»). Другой пример: вы подбрасываете кубик, на разных гранях которого изображены различные числа очков, и кубик падает так, что сверху оказывается грань с четырьмя очками. В данном случае второе событие (выпала четверка) уже не является однозначным исходом первого события (подброшен кубик). Ведь могли выпасть единица, двойка, тройка, пятерка, шестерка. Выпадение того или иного числа очков есть пример *случайного* события. Из приведенных примеров хорошо видно различие между строго детерминированными и случайными событиями.

Со случайными событиями (и вообще со случайностями разного рода) мы встречаемся очень часто, значительно чаще, чем это обычно принято считать. Случаен набор выигравших номеров в тираже спортлото. Случаен результат встречи двух спортивных команд одного и того же класса. Количество солнечных дней в данной местности изменяется от года к году случайным образом. Совокупность случайных факторов лежит в основе любого процесса массового обслуживания — телефонной связи, торговли, транспортных услуг, медицинской помощи и т. д.

В интересной книге «Вероятность в играх и развлечениях», написанной французским педагогом Морисом Глеманом и венгерским педагогом Тамашом Варгой (М.: Просвещение, 1979), есть очень глубокое замечание: «Сталкиваясь со случайной ситуацией, маленькие дети думают, что можно *предсказать* ее исход; становясь немного постарше, они отвечают, что *ничего нельзя утверждать*; но мало-помалу они открывают, что за кажущимся хаосом мира случайности можно обнаружить законы, которые позволяют неплохо ориентироваться в реальности». Здесь выделены три после-

довательные стадии — сначала просто непонимание случайности, затем растерянность и, наконец, правильная точка зрения. Забудем на время о маленьких детях и попробуем «примерить» это замечание к самим себе. Нам придется признать, что довольно часто мы останавливаемся на первой стадии, наивно полагая, будто в конечном счете всякий исход может быть точно предсказан. До сих пор пользуется популярностью возникшее много столетий назад заблуждение, когда случай попросту приравнивают к хаосу, к отсутствию причинности. И сегодня далеко не все достаточно ясно представляют себе, что за обилием окружающих нас случайностей скрыты специфические закономерности (*вероятностные* закономерности).

Все это и побудило автора написать данную книгу. Автор стремился помочь читателю открыть для себя вероятностную природу окружающего мира, познакомить со случайными явлениями, показать, что в мире случайностей можно хорошо ориентироваться и, более того, активно действовать.

Книга открывается беседой автора с воображаемым читателем о роли случайности, а завершается беседой о взаимосвязи случайности и симметрии. Основной материал книги разбит на две части. В первой части обсуждается понятие вероятности и рассматриваются различные применения вероятности в человеческой практике — для принятия решений в сложных ситуациях, при организации массового обслуживания, при проведении игр, для оптимизации управления различными процессами, при случайном поиске и т. д. Даются представления о кибернетике и теории информации, о таких новых научных направлениях, как исследование операций и теория игр. Познакомившись с первой частью книги, читатель убедится, что мир случайностей начинается сразу же за порогом его комнаты, поскольку фактически вся современная человеческая деятельность опирается на вероятностные методы. Во второй части книги раскрывается фундаментальная роль случайности в природе — на примере вероятностных законов современной физики и биологии. Здесь же рассматриваются элементы квантовой механики, что позволяет продемонстрировать фундаментальность вероятностных законов на уровне микроявлений. Перейдя от первой части книги ко второй, читатель, по замыслу автора, должен убедиться, что вероятность не только вокруг нас, но и в основе всего.

В заключение автор хотел бы выразить признательность всем, кто участвовал в создании данной книги. Идея ее написания была подсказана автору членом-корреспондентом АН СССР И. И. Гуревичем; он же высказал ряд интересных мыслей, касающихся отбора материала и структуры книги. С рукописью книги внимательно ознакомились и сделали ценные замечания академик Б. В. Гнеденко, доктор философских наук Г. Я. Мякишев, кандидат физико-математических наук О. Ф. Кабардин. В процессе работы над рукописью автор неизменно получал поддержку и полезные советы от В. А. Ежова и А. Н. Тарасовой.

Введение

О, сколько нам открытий чудных
Готовят просвещенья дух
И опыт, сын ошнбок трудных,
И гений, парадоксов друг,
И случай, бог-нзобретатель...

А. С. Пушкин

Разговор автора с читателем о роли случайности

Читатель. В предисловии сказаио немало похвальных слов случаю. Несмотря на это, мие все же кажется, что в целом случайность нграет отрицательиую роль. Конечио, случай может оказаться счастливым. Но, как известио, рассчитывать иа иего ие рекомеидуется. Случайиости нам мешают, путают иашн плаиы. Лучше от иих ие зависеть н стараться, по возможиостн, исключить их из иашей практики.

Автор. Именно таково траднцноиное отношение к случайному. Одиако в нашн дни оно явно нуждается в пересмотре. Прежде всего, действительно лн возможио исключнть случайиое из иашей практики?

Читатель. Я ие утверждал, что это можио сделать. Я говорил лишь, что иадо стараться это делать.

Автор. Предположнм, что вы работаете иа стаиции скорой медициской помощи. Ясио, что вы не можете предвидеть, когда имеиио потребуется срочиая помощь, куда надо будет ехать, сколько времени придется затратить иа того или иного больного. А ведь от всего этого зависит ответ иа практическнй вопрос: сколько дежуриых врачей иадо иметь, чтобы, с одиой стороиы, им ие приходилось долго бездействовать в ожндании вызова, а с другой — больным ие приходилось слишком долго ждать помощи? Исключить случайиости вы ие можете. Поэтому для получения ответа иа поставлеиный вопрос вы должны постараться наилучшим образом учесть их. Подчеркиваю: не *исключить*, а *учесть*.

Читатель. Действительно, здесь приходнтся мириться со случайиостями. Одиако оии продолжают оставаться отрицательным фактором.

Автор. Итак, мы убеднлнсь, что ннogда прнходнтся переходить от борьбы со случайностямн к союзу с ними. Но можио пойти и дальше. Можно указать ситуацин, когда случайности из отрицательиого фактора превращаются в положнтельный, так что желательно специально повышать уровень случайиого.

Чнтатель. Этого я ие понимаю.

Автор. Конечно, случайностн путают наши планы. Но ведь в то же время они заставляют иас искать новые решения, развивают иашу способиость к актнвной деятельностн.

Читатель. Развнтие в процессе преодолеиия трудностей?

Автор. Главное в том, что случайности могут создавать новые возможности. У американского писателя Р. Ф. Джоунса есть любопытная фантастическая повесть «Уровень шума» (см. 10-й том «Библиотеки современной фантастики»). Группе ученых, являющихся специалистами в разных областях, официально сообщают, что сделано сенсационное открытие, но, к несчастью, изобретатель погиб при взрыве во время показа своего изобретения и унес его тайну с собой. В действительности не было ни изобретения, ни погибшего изобретателя. Ученым предъявляют все, что якобы осталось после гибели изобретателя: неразборчивые обрывки записей, лабораторию с множеством различных приборов и устройств, библиотеку. Одним словом, ученым предложена обширная несистематизированная информация, насыщенная случайными сведениями из различных областей науки и техники; ее можно назвать информационным шумом. И вот, будучи уверены, что изобретение состоялось на самом деле, и, следовательно, поставленная задача имеет решение, ученые успешно используют предоставленный в их распоряжение информационный шум и раскрывают тайну несуществовавшего изобретения. Можно сказать, что им удается успешно произвести отбор информации из шума.

Читатель. Но ведь это всего лишь фантастическая повесть.

Автор. Конечно. Однако идея повести отнюдь не фантастична. Всякое открытие связано с использованием случайных факторов.

Читатель. Не думаю, чтобы без глубоких специальных знаний можно было бы чисто случайно открыть что-либо серьезное.

Автор. Согласен с вами. Более того, для открытия требуется не только определенная квалификация исследователя, но и уровень развития науки в целом. И все же фактор случайного играет при этом фундаментальную роль.

Читатель. Как я понимаю, фундаментальное — это нечто первичное, то, что лежит в основе. Можно ли применять термин «фундаментальный» к случайному? Я допускаю, что случайность может быть полезной. Но может ли она быть фундаментальной? Ведь в конечном счете мы имеем дело со случайностями тогда, когда чего-то не знаем, что-то не можем учесть.

Автор. Полагая, что случайность связана с неполнотой наших знаний, вы тем самым относите ее к понятиям *субъективным*. Получается, что случайное лежит как бы на поверхности, тогда как в глубине, в основе явлений случайного нет. Не так ли?

Читатель. Именно так. Поэтому и нельзя говорить о фундаментальности случайного. По мере развития науки расширяются наши возможности учитывать различные факторы, и в результате область проявления случайного должна постепенно сужаться. Недаром же говорят, что «наука — враг случайностей».

Автор. Вы не совсем правы. Развитие науки действительно увеличивает возможности научных предсказаний, т. е. действует против фактора случайного. В то же время оказывается, что по мере углубления научных знаний, а точнее, при переходе на молекулярный и атомный уровни рассмотрения явлений случайность в

целом не только не уменьшается, но, напротив, начинает по-настоящему господствовать. Ее существование оказывается уже не зависящим от степени наших знаний. Именно на уровне микромира случайность и обнаруживает свою *фундаментальность*.

Читатель. Я впервые слышу об этом. Поясните свою мысль.

Автор. Предварительно я хотел бы заметить, что у данного вопроса длинная история. Фактически она началась во времена античной философии, когда наметились два подхода к случайному. Один связан с именем Демокрита, а другой Эпикура. Демокрит отождествлял случайное с *непознанным*, полагая, что в своей основе природа строго детерминирована Он говорил: «Люди сотворили себе кумира из случая как прикрытие для присущего им недомыслия». Эпикур же считал, что случай присущ самой природе явлений и что, следовательно, случайность *объективна*. Весьма долгое время предпочтение отдавалось точке зрения Демокрита. И лишь в XX веке развитие науки показало, что более верна точка зрения Эпикура. Карл Маркс в своей докторской диссертации «Различие между натурфилософией Демокрита и натурфилософией Эпикура» положительно оценивал взгляды Эпикура на случайное, указывал на глубокое философское значение выдвинутого Эпикуром учения о *спонтанном* (самопроизвольном) «отклонении атомов». Разумеется, не следует преувеличивать вклад Эпикура в учение о случайном, это были всего лишь догадки.

Читатель. Получается, что, сам того не ведая, я излагал взгляды Демокрита на случайность. Но хотелось бы познакомиться с конкретными примерами, показывающими фундаментальность случайного.

Автор. Рассмотрим, например, атомную подводную лодку. Ее двигатель работает на основе ядерного реактора. Как привести этот двигатель в действие?

Читатель. Насколько я знаю, надо начать выводить из активной зоны реактора специальные стержни, предназначенные для поглощения нейтронов. Тогда начнется управляемая цепная реакция деления ядер урана...

Автор (перебивая). Попробуем проследить за тем, как она начинается.

Читатель. Попадая в ядро урана, нейтрон вызывает его деление на два осколка. При этом освобождается некоторая энергия и появляются два свободных нейтрона. Новые нейтроны вызовут деление уже двух ядер урана; при этом появятся четыре нейтрона, которые, в свою очередь, вызовут деление уже четырех ядер. Процесс развивается подобно лавине.

Автор. Хорошо. Но откуда берется первый нейтрон?

Читатель. Мало ли откуда... Хотя бы из космического излучения.

Автор. Подводная лодка находится на значительной глубине. Толстый слой воды хорошо защищает ее от космического излучения.

Читатель. В таком случае не знаю...

Автор. Дело в том, что ядро урана может разделиться не только при попадании в него нейтрона, но и *самопроизвольно*, иначе говоря, спонтанно. *Процесс спонтанного деления ядра случаен.*

Читатель. Но, может быть, спонтанное деление ядер в конечном счете обусловлено факторами, которые мы пока еще не знаем?

Автор. Этот вопрос физики задавали себе много раз. Предпринимались попытки отыскать так называемые «скрытые параметры», которые могли бы рассматриваться как факторы, управляющие процессами в микромире. В итоге пришли к выводу, что таковых параметров нет и что, следовательно, случайность в явлениях микромира действительно фундаментальна. Этот принципиальный вопрос рассматривается в *квантовой механике* — физической теории, которая возникла в первой половине XX века в связи с исследованиями атомных процессов.

Читатель. О квантовой механике я знаю только то, что она описывает законы движения микрочастиц.

Автор. Позднее мы поговорим о квантовой механике более подробно. Пока же отметим, что она показала принципиальную роль спонтанных процессов и тем самым продемонстрировала фундаментальность случайного. Забегая вперед, заметим, что без спонтанных процессов была бы невозможна работа любого генератора излучения — от обычного лампового генератора до лазера. Эти процессы исполняют фундаментальную роль «начального сигнала», без которого не могло бы начаться развитие процесса генерации.

Читатель. И все же мне трудно принять идею фундаментальности случайного. Вы приводили пример с атомной подводной лодкой. Но ведь командир лодки, подавая команду включить двигатели, не рассчитывает на случайность. Нажимается соответствующая кнопка — и двигатели обязательно начинают работать (при условии, что они исправны). То же можно сказать и о включении лампового генератора. Где же тут случайность?

Автор. И тем не менее на уровне явлений микромира процессы в рассматриваемых устройствах начинаются от случайных факторов.

Читатель. На практике, однако, мы имеем дело с процессами, происходящими в макромире.

Автор. Во-первых, исследуя окружающий мир, стремясь понять его *причинно-следственные связи*, мы неизбежно выходим на атомный уровень, т. е. на уровень явлений микромира. Во-вторых, случайное в явлениях микромира может существенно отражаться на характере процессов, наблюдаемых в масштабе макромира.

Читатель. А нельзя ли привести пример, когда фундаментальность случайного обнаруживается в масштабе макромира?

Автор. Таким примером может служить *эволюция*, непрерывно совершающаяся в растительном и животном мире. В основе эволюции лежат *мутации* — случайные изменения в структуре генов. Случайно возникшая мутация способна быстро усилиться в процессе размножения клеток организма. Существенно, что одновременно с мутациями (случайными изменениями генетических программ) происходит процесс *отбора* организмов. Отбор совершается по сте-

пени приспособленности к условиям внешней среды. Таким образом, эволюция основывается на *отборе случайных изменений генетических программ.*

Читатель. Не совсем понятно, как именно действует отбор.

Автор. Рассмотрим пример. У некоторых орхидей цветы напоминают самок шмелей. Опыляются они самцами шмелей, которые принимают цветы за самок. Предположим, что возникла мутация, изменившая форму или окраску цветка. Такой цветок останется неопыленным. В результате мутация не перейдет в новое поколение. Можно сказать, что отбор забраковал мутацию, изменившую внешний вид цветка. Любопытно, что, когда один из видов орхидей стал самоопылителем, цветы этого вида быстро приобрели за счет мутаций разнообразную форму и окраску.

Читатель. Насколько я знаю, эволюция идет в направлении усложнения видов. Не указывает ли это на то, что лежащие в основе эволюции мутации в действительности не так уж случайны?

Автор. Вы не правы. Эволюция идет не по пути отбора более сложных, а по пути отбора *более приспособленных* организмов. А на этом пути иногда предпочтительна более высокая степень организации, а иногда — наоборот. Недаром же в современном мире существуют одновременно и человек, и медуза, и вирус гриппа. Существенно, что эволюция приводит к появлению принципиально непредсказуемых новых видов. Можно утверждать, что *любой вид уникален, ибо он принципиально случаен.*

Читатель. Надо признать, что здесь случайность действительно выглядит как фундаментальный фактор.

Автор. Говоря о фундаментальности случайного в картине эволюции, отметим еще одно немаловажное обстоятельство. Понимание фундаментальной роли случайного позволяет отбросить религиозную идею о сверхъестественном «творце». Служители церкви, отвечая на вопрос, как возникли растения, животные, человек, указывают на бога. Образованный же человек должен понимать, что вместо несуществующего бога в роли «творца» выступает *случай* и *отбор.*

Читатель. Прямо по Пушкину: «и случай, бог-изобретатель...»

Автор. Именно так. Поразительно, как точно выразился поэт.

Читатель. Говоря о случае и отборе, следует, по-видимому, подразумевать *отбор информации из шума?* Тот самый, о котором шла речь при обсуждении повести «Уровень шума».

Автор. Совершенно верно.

Читатель. Приходится согласиться с тем, что следует не столько бороться со случайностями, сколько сознательно идти навстречу им.

Автор. Следовало бы выразиться точнее. Случайность, связанная с неполнотой наших знаний, конечно, нежелательна. Познавая окружающий мир, человек боролся, борется и всегда будет бороться с ней. В то же время необходимо понимать, что наряду с *субъективной* случайностью, обусловленной недостатком сведений о тех или иных явлениях, существует *объективная* случайность, лежащая в самой основе явлений. Необходимо также

принимать во внимание позитивную, созидательную роль случайного. И в этой связи действительно надо идти навстречу случайности. Человек должен уметь, когда это целесообразно, специально создавать ситуации, насыщенные случайностями, и использовать подобные ситуации в своих целях.

Читатель. Но можно ли в принципе обращаться со случайностью таким образом? Не похоже ли это на *управление неуправляемым?*

Автор. Наука и практика показали, что сознательно ориентироваться в ситуациях, насыщенных случайностями, можно. Разработаны специальные расчетные методы, использующие фактор' случайного. Созданы специальные теории, такие, например, как *теория массового обслуживания, теория игр, теория случайного поиска,* и другие.

Читатель. Мне трудно представить себе научную теорию, построенную на случайности.

Автор. Сразу же подчеркнем: случайное отнюдь не исключает возможности научных предсказаний. Из фундаментальности случайного вовсе не следует вывод о беспорядочности и хаотичности окружающего нас мира. Случайность совсем не означает, что причинно-следственные связи отсутствуют. Но обо всем этом позже. А пока попробуйте представить себе мир, в котором случайность как объективный фактор *полностью отсутствует.*

Читатель. Это был бы идеально упорядоченный мир.

Автор. В таком мире состояние любого объекта в данный момент времени однозначно определялось бы его прошлыми состояниями и, в свою очередь, столь же однозначно определяло бы будущие состояния. Прошлое было бы жестко связано с настоящим, а настоящее с будущим.

Читатель. Все происходящее в подобном мире было бы заранее предопределено.

Автор. Известный французский ученый XVII века П. Лаплас предлагал в связи с этим вообразить некое «сверхсущество», которому досконально известно прошлое и будущее такого мира. «Ум, которому были бы известны для какого-либо момента времени все силы, одушевляющие природу, и относительные положения всех ее составных частей,— писал Лаплас,— если бы вдобавок он оказался достаточно обширным, чтобы подвергнуть эти данные анализу, обнял бы в одной формуле движения величайших тел Вселенной наравне с движениями легчайших атомов. Не осталось бы ничего, что было бы для него недостоверно, и будущее, так же как и прошедшее, предстало бы перед его взором».

Читатель. Идеально упорядоченный мир оказывается нереальным.

Автор. Как видите, совсем нетрудно почувствовать, что реальный мир должен допускать существование объективной случайности. Теперь вернемся к вопросу о причинно-следственных связях. В реальном мире эти связи являются *вероятностными.* Лишь в отдельных случаях (в частности, при решении задач из школьных

задачников) мы имеем дело с однозначными, строго детерминированными связями. Здесь мы подошли к одному из важнейших понятий современной науки — понятию *вероятности*.

Читатель. Оно мне знакомо. Если, например, я бросаю игральных кубик, то с одинаковым успехом могу ожидать выпадения любого из шести чисел. Можно утверждать, что вероятность выпадения того или иного числа одна и та же; она равна 1/6.

Автор. А какова вероятность того, что у четырехзначного номера случайно проезжающего мимо вас автомобиля (допустим, что вы стоите на обочине шоссе) две первые цифры одинаковы?

Читатель. Эта вероятность равна 1/10.

Автор. Значит, если вы наберетесь терпения и понаблюдаете за достаточно большим числом автомобилей, то примерно десятая часть их будет иметь номера с двумя одинаковыми цифрами? Скажем, из 300 машин такие номера будут примерно у 30. Может быть, у 27 или 32, но никак не у 10 или 100.

Читатель. По-видимому, так и будет.

Автор. Но в таком случае вам незачем стоять на обочине шоссе. Результат может быть предсказан заранее. Это и есть пример *вероятного предсказания*. Обратите внимание на обилие случайных факторов в данной ситуации. Какая-то машина могла свернуть с шоссе, не доезжая, какая-то могла остановиться или даже повернуть назад. И тем не менее и сегодня, и завтра из 300 машин примерно 30 будут иметь номера с одинаковыми двумя первыми цифрами.

Читатель. Получается, что, несмотря на множество случайных факторов данная ситуация обнаруживает некое постоянство.

Автор. Это справедливо подмеченное постоянство обычно называют *статистической устойчивостью*. Существенно, что статистическая устойчивость наблюдается не «вопреки случайным факторам», а благодаря наличию этих факторов.

Читатель. Я как-то не задумывался над тем, что сплошь и рядом мы имеем дело с вероятностными предсказаниями. Ведь к ним относятся, например, спортивные прогнозы, предсказания погоды.

Автор. Вы совершенно правы.

Важно подчеркнуть, что вероятностные (статистические) причинно-следственные связи являются общим видом связей, тогда как связи, приводящие к однозначным предсказаниям, представляют собой всего лишь частный случай. Если однозначные предсказания предполагают наличие в рассматриваемом явлении только необходимости, то вероятностные предсказания связаны одновременно и с необходимостью, и со случайностью. Так, мутации случайны, но процесс отбора закономерен или, иначе говоря, необходим.

Читатель. Понимаю. Отдельные акты спонтанного деления ядер урана случайны, но развитие цепной реакции необходимо.

Автор. Само по себе то или иное открытие случайно. Однако должна существовать ситуация, благоприятствующая возникнове-

нию такой случайности. Она определяется развитием науки, уровнем измерительной техники, квалификацией исследователей. Открытие случайно, но необходима (закономерна) логика развития, приводящая в конечном счете к открытию.

Читатель. Мне понятно теперь, почему фундаментальность случайного не приводит к беспорядочности нашего мира. Случайное и необходимое всегда выступают вместе.

Автор. Правильно. Как писал Фридрих Энгельс, «в природе, где как будто господствует случайность, мы давно уже установили в каждой отдельной области внутреннюю необходимость и закономерность, которые пробивают себе дорогу в рамках этой случайности»[1]. Об этом же написал в своей интересной книге «Письма о вероятности» венгерский математик А. Реньи: «...Я наткнулся на «Размышления» Марка Аврелия и случайно открыл ту страницу, где он пишет о двух возможностях: либо мир является огромным хаосом, либо в нем царствуют порядок и закономерность. И хотя я уже много раз читал эти строки, но теперь впервые задумался над тем, а почему, собственно, Марк Аврелий считал, что в мире господствуют либо случайность, либо порядок и закономерность? Почему он думал, что эти две возможности исключают друг друга? В мире господствует случай и одновременно действуют порядок и закономерность, которые формируются из массы случайностей согласно законам случайного».

Читатель. Насколько я понимаю, формирование порядка и закономерности из массы случайностей и приводит к понятию вероятности?

Автор. Совершенно верно. Обратите внимание: *отдельные элементы* меняются от случая к случаю. В то же время *картина в целом* обнаруживает устойчивость. Эта устойчивость и выражается через *вероятность*. Поэтому наш мир оказывается достаточно гибким, динамичным, способным к развитию.

Читатель. Отсюда следует, что окружающий нас мир можно с полным основанием назвать вероятностным миром.

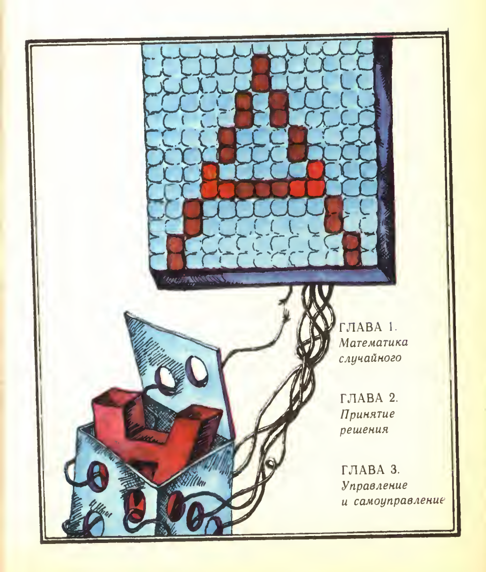
Автор. Лучше говорить *о мире, построенном на вероятности*. Переходя к исследованию этого мира, мы сосредоточим внимание на двух группах вопросов. Во-первых, покажем, как за счет широкого применения вероятности в своей научной и практической деятельности человек сумел «приручить» случайность и тем самым превратил ее из извечного противника в союзника и помощника. Во-вторых, используя достижения современной физики и биологии, продемонстрируем вероятностный характер фундаментальных законов природы. Тем самым наглядно покажем, что весь окружающий мир (включая как природу, так и тот мир, который создает человек в процессе своей деятельности) действительно построен на вероятности.

---

[1] Маркс К., Энгельс Ф. Происхождение семьи, частной собственности и государства.— Соч., 2-е изд., т. 21, с. 174.

Часть I

Прирученная случайность



ГЛАВА 1 Математика случайного

Это учение, объединяющее точность математических доказательств с неопределенностью случая и примиряющее эти, казалось бы, противоречивые элементы, с полным правом может претендовать на титул — математика случайного.

Блез Паскаль

Вероятность

Классическое определение вероятности. Когда мы подбрасываем монету, мы не знаем, что именно выпадет — герб или «решка». Однако кое-что мы все же знаем. Мы знаем, что шансы выпадения как герба, так и «решки» одинаковы. Точно так же мы знаем, что одинаковы шансы выпадения любой из шести граней игрального кубика. В обоих примерах равенство шансов связано с *симметрией.* Симметрична монета, симметричен кубик. Будем называть *равновозможными* исходы, имеющие одинаковые шансы. Выпадение герба и выпадение «решки» — равновозможные исходы.
Предположим, что нас интересует определенный результат бросания игрального кубика, например, выпадение грани с числом очков, делящимся без остатка на три. Будем называть *благоприятными* исходы, при которых получается этот результат. В данном случае имеем два таких исхода — выпадение тройки и выпадение шестерки. Наконец, будем называть исходы *несовместными,* если при появлении одного из них в единичном испытании исключается появление другого в том же испытании. Выпадения граней при бросании кубика — несовместные исходы.
Теперь мы можем сформулировать классическое определение вероятности. *Вероятностью события называется отношение числа благоприятных исходов к общему числу несовместных равновозможных исходов.*
Пусть P_A — вероятность события A, m_A — число благоприятных исходов, n — общее число несовместных равновозможных исходов. Согласно классическому определению вероятности

$$P_A = m_A/n. \qquad (1.1)$$

Если $m_A = n$, то $P_A = 1$. Событие A есть *достоверное* событие (оно реализуется в каждом исходе). Если $m_A = 0$, то $P_A = 0$. Событие A есть *невозможное* событие (оно вообще не реализуется). Вероятность *случайного* события лежит между 0 и 1.
Пусть событие A — выпадение грани кубика с числом очков, делящимся без остатка на три. В этом случае $m_A = 2$. Поскольку $n = 6$, то вероятность данного события есть 1/3. Рассмотрим еще один пример. В мешке находятся 15 шаров, различающихся только по цвету (7 белых, 2 зеленых и 6 красных). Вы вытаскиваете на-

угад один шар. Какова вероятность того, что извлеченный из мешка шар окажется белым (красным, зеленым)? Извлечение белого шара будем рассматривать как событие A, красного — как событие B, зеленого — как событие C. Число исходов, благоприятных для извлечения шара того или иного цвета, равно числу шаров соответствующего цвета: $m_A=7$, $m_B=6$, $m_C=2$. Используя формулу (1.1) и учитывая, что $n=15$, находим искомые вероятности:

$$P_A=\frac{m_A}{n}=\frac{7}{15}; \qquad P_B=\frac{m_B}{n}=\frac{2}{5}; \qquad P_C=\frac{m_C}{n}=\frac{2}{15}.$$

Сложение и умножение вероятностей. Какова вероятность того, что наугад извлеченный шар окажется либо красным, либо зеленым? Число благоприятных исходов $m_B+m_C=6+2=8$, поэтому искомая вероятность равна $P_{B+C}=(m_B+m_C)/n=8/15$. Мы видим, что $P_{B+C}=P_B+P_C$. Вероятность вытащить либо красный, либо зеленый шар равна сумме двух вероятностей: вероятности вытащить красный и вероятности вытащить зеленый. Вероятность вытащить шар, цвет которого будет либо красным, либо зеленым, либо белым, есть сумма трех вероятностей: $P_A+P_B+P_C$. Она равна единице $(7/15+2/5+2/15=1)$. Это естественно, поскольку рассматриваемая вероятность есть вероятность достоверного события.

Правило сложения вероятностей может быть сформулировано следующим образом. *Вероятность того, что произойдет какое-либо из нескольких несовместных событий, равна сумме вероятностей рассматриваемых событий.*

Предположим, что подбрасываются одновременно два кубика. Какова вероятность того, что одновременно выпадут две четверки? Общее число несовместных равновозможных исходов $n=6\times6=36$. Все они отмечены на рисунке 1.1, где левая цифра в скобках — очки одного кубика, а правая — очки другого. Имеется только один благоприятный исход, он отмечен на рисунке как (4;4). Следовательно, искомая вероятность равна 1/36. Эта вероятность есть произведение двух вероятностей: вероятности выпадения четверки для одного кубика и вероятности выпадения четверки для другого кубика:

$$P_{44}=P_4\times P_4=\frac{1}{6}\times\frac{1}{6}=\frac{1}{36}.$$

Правило умножения вероятностей формулируется так: *вероятность того, что произойдут сразу несколько событий, равна произведению вероятностей этих событий.*

Впрочем, буквально понимаемая одновременность событий необязательна. Вместо того чтобы подбрасывать одновременно два кубика, можно подбросить два раза один и тот же кубик. Вероятность одновременного выпадения четверок при подбрасывании двух

(1;1)	(2;1)	(3;1)	(4;1)	(5;1)	(6;1)
(1;2)	(2;2)	(3;2)	(4;2)	(5;2)	(6;2)
(1;3)	(2;3)	(3;3)	(4;3)	(5;3)	(6;3)
(1;4)	(2;4)	(3;4)	(4;4)	(5;4)	(6;4)
(1;5)	(2;5)	(3;5)	(4;5)	(5;5)	(6;5)
(1;6)	(2;6)	(3;6)	(4;6)	(5;6)	(6;6)

Рис. 1. 1

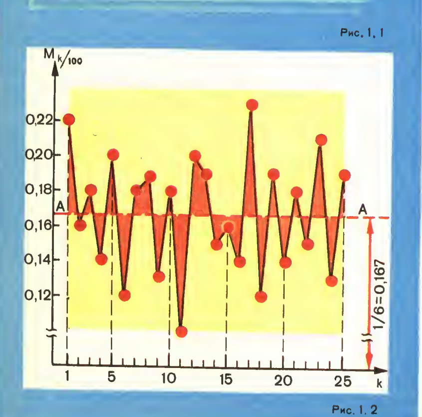


Рис. 1. 2

кубиков совпадает с вероятностью того, что четверка выпадет при обоих подбрасываниях одного и того же кубика.

Во многих случаях при вычислении вероятности события применяют совместно оба правила (сложения и умножения вероятностей). Пусть нас интересует вероятность P того, что при подбрасывании двух кубиков выпадут *одинаковые* очки на обоих кубиках. Поскольку безразлично, какие именно очки выпадут (важно, чтобы они были одинаковыми), то можно воспользоваться правилом сложения вероятностей:

$$P=P_{11}+P_{22}+P_{33}+P_{44}+P_{55}+P_{66}.$$

В свою очередь, каждая из вероятностей P_{ii} есть произведение $P_i \times P_i$. Таким образом,

$$P=(P_1\times P_1)+(P_2\times P_2)+\dots+(P_6\times P_6)=6\left(\frac{1}{6}\times\frac{1}{6}\right)=\frac{1}{6}.$$

Этот результат можно получить сразу из рисунка 1.1, где благоприятные исходы выделены красным цветом: (1; 1), (2; 2), (3; 3), (4; 4), (5; 5), (6; 6). Всего таких исходов шесть. Следовательно, $P=6/36=1/6$.

Частота и вероятность. Исходя из классического определения вероятности и применяя правила сложения и умножения вероятностей, мы можем рассчитать вероятность того или иного случайного события. Какова, однако, практическая ценность подобных расчетов? Что, например, означает на практике утверждение, что вероятность выпадения четверки при подбрасывании кубика равна 1/6? Разумеется, это утверждение не означает, что при шести бросаниях четверка должна выпасть один, и только один раз. Возможно, что она выпадет один раз, но возможно также, что она выпадет два (и более) раза или же не выпадет совсем. Чтобы проявилась вероятность, надо проделать большое число бросаний и проследить, насколько часто выпадает четверка.

Выполним несколько серий бросаний кубика, делая в каждой серии, скажем, 100 бросаний. Обозначим числа выпадения четверки в 1-й, 2-й, 3-й и т. д. сериях соответственно через M_1, M_2, M_3 ... Отношения $M_1/100$, $M_2/100$, $M_3/100$ ... представляют собой частоты появления четверки в соответствующих сериях. Выполнив некоторое количество серий бросаний, можно убедиться, что частота появления четверки от серии к серии *случайным образом колеблется около вероятности* данного события, т. е. около 1/6. Это видно на рисунке 1.2, где по оси абсцисс отложены номера k серий, а по оси ординат полученные в некоем эксперименте значения частоты появления четверки. Разумеется, если заново повторить подобный эксперимент, то получатся иные значения частот $M_k/100$. Однако картина колебаний частот появления рассматриваемого события обнаружит устойчивость: отклонения вверх и вниз от прямой AA, отвечающей вероятности события, будут взаимно уравновешены, амплитуда отклонений хотя и будет меняться от серии к серии, в то же время не проявит тенденции ни к возрастанию, ни к за-

туханию. Это есть следствие *равноправия* серий. Ведь число испытаний в каждой серии одинаково, и, кроме того, результат испытаний, полученный в данной серии, не зависит от результата испытаний в предыдущей серии.

Сделаем важный шаг: перейдем к сериям с *постепенно возрастающим* числом испытаний. Используя результаты эксперимента, представленные на рисунке 1.2, будем рассматривать новые серии, которые получаются при *объединении* двух первых прежних серий, трех первых, четырех первых и т. д. Иначе говоря, учтем число появлений четверки сначала при первых 100 бросаниях (в нашем случае $M_1=22$), затем при первых 200 бросаниях ($M_1+M_2=$ $=22+16=38$), при первых 300 бросаниях ($M_1+M_2+M_3=$ $=22+16+18=56$) и т. д. Выпишем частоты появления четверки в новых сериях: $M_1/100=0{,}22$; $(M_1+M_2)/200=0{,}19$; $(M_1+M_2+$ $+M_3)/300=0{,}187$... На рисунке 1.3 представлены эти частоты, рассмотрены последовательно серии с числом бросаний 100, 200, ..., 2500. Рисунок демонстрирует весьма важный факт: отклонение частоты появления события от его вероятности уменьшается по мере увеличения числа испытаний. Иными словами, *при увеличении числа испытаний частота появления случайного события приближается к его вероятности.*

Возможно ли частотное определение вероятности? Поскольку при увеличении числа испытаний частота появления события приближается к его вероятности, возникает вопрос: нельзя ли определить вероятность события как предел отношения числа его появлений к числу испытаний, вычисленный при неограниченном возрастании числа испытаний? Пусть N — число испытаний, а $M_A(N)$ число появлений события A в этих испытаниях. Спрашивается, нельзя ли определить вероятность P_A события A следующим образом:

$$P_A=\lim_{N\to\infty}[M_1(N)/N]. \qquad (1.2)$$

Немецкий математик первой половины XX века Р. Мизес полагал, что выражение (1.2) можно рассматривать как определение вероятности случайного события, он называл его *частотным* определением вероятности. Мизес указывал, что классическое определение вероятности (1.1) «работает» лишь тогда, когда имеется *конечное* число *равновозможных* исходов. Таковы, например, ситуации, связанные с подбрасыванием монеты или игрального кубика.

На практике мы часто встречаемся с ситуациями, где нет *симметрии,* предопределяющей равновозможность исходов. В таких случаях классическим определением вероятности пользоваться нельзя. Вот здесь, как утверждал Мизес, и может пригодиться частотное определение, поскольку оно не нуждается в конечном числе равновозможных исходов и, более того, вообще не предполагает вычисления вероятности. При частотном подходе вероятность не вычисляется, а определяется из опыта.

0,22
0,21
0,20
0,19
0,18
0,17
0,16
A
A
1/6 =0,167
N
100 500 1000 1500 2000 2500

Рис.1.3

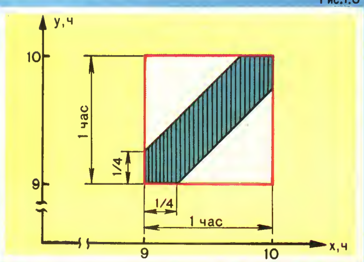


Рис.1.4

Можно ли, однако, на практике определить вероятность какого-либо случайного события, используя соотношение (1.2)? Это соотношение предполагает, что выполняется *бесконечно большое* число однотипных испытаний. На практике пришлось бы ограничиться *конечным* числом испытаний. При этом совершенно неясно, каким именно оно должно быть. Можно ли ограничиться сотней испытаний? Или надо выполнить тысячу, миллион, сто миллионов испытаний? И с какой точностью при этом определяется искомая вероятность? На все эти вопросы ответа не существует. К тому же на практике невозможно обеспечить одинаковые условия выполнения очень большого числа испытаний. Мы уже не говорим о том, что сам характер испытаний может сделать невозможным их многократное повторение.

Таким образом, соотношение (1.2) оказывается практически бесполезным. Более того, можно показать (мы этого делать не будем), что фигурирующий в (1.2) предел, строго говоря, *не существует*. Это означает, что предложенное Мизесом соотношение (1.2) не только практически бесполезно, но и не имеет смысла. Следовательно, оно не может рассматриваться в качестве определения вероятности. Иными словами, частотное определение вероятности оказывается несостоятельным. Ошибка Мизеса заключалась в том, что исходя из правильной посылки (при увеличении числа испытаний частота появления случайного события приближается к его вероятности), он сделал необоснованный вывод, будто вероятность события есть предел частоты его появления при неограниченном возрастании числа испытаний.

Геометрическое определение вероятности. Предположим, что два человека условились о встрече в некотором месте между девятью и десятью часами. Они договорились, что каждый ждет другого в течение четверти часа, а затем уходит. Какова вероятность, что они встретятся? Пусть x — момент прихода одного человека на место встречи, а y — момент прихода другого. Точку на плоскости с координатами (x, y) будем рассматривать как один из исходов встречи. Все возможные исходы лежат в площади квадрата, сторона которого соответствует промежутку времени длительностью в один час (рис. 1.4). Исход будет благоприятным (встреча состоится), если точка (x, y) такова, что $|x - y| \leqslant 1/4$. Такие точки лежат в пределах заштрихованного на рисунке участка площади квадрата. Все исходы равновозможны и несовместны, поэтому вероятность встречи равна отношению заштрихованной площади ко всей площади квадрата.

Это напоминает использовавшееся в классическом определении вероятности отношение числа благоприятных исходов к общему числу равновозможных исходов. Следует иметь в виду, что в данном случае число исходов (как всех, так и благоприятных) бесконечно. Поэтому здесь говорят не об отношении чисел соответствующих исходов, а об отношении площади области, благоприятствующей появлению рассматриваемого случайного события, к площади всей области.

Используя рисунок 1.4, нетрудно найти площадь благоприятной области. Она равна разности площадей всего квадрата и его незаштрихованной части: $1-(3/4)^2=7/16$ ч^2. Разделив $7/16$ ч^2 на 1 ч^2, получаем вероятность встречи: 7/16.

Рассмотренный пример иллюстрирует *геометрическое* определение вероятности: *вероятность случайного события есть отношение площади области, благоприятствующей появлению события, к площади всей области.* Геометрическое определение вероятности является обобщением классического определения на случай, когда число равновозможных исходов бесконечно.

Развитие понятия вероятности. Вероятностные представления достаточно широко использовались уже древнегреческими философами (Демокрит, Эпикур, Лукреций Кар и др.). Как наука теория вероятностей стала развиваться лишь с середины XVII века — в работах французских ученых Б. Паскаля и П. Ферма, а также голландского ученого X. Гюйгенса. Классическое определение вероятности случайного события было сформулировано в знаменитом труде «Наука предположений» известного швейцарского математика Я. Бернулли. Окончательно это определение оформилось позднее — в работах П. Лапласа. Геометрическое определение вероятности стали применять в XVIII веке. Существенный вклад в развитие теории вероятностей внесла русская математическая школа в XIX веке (П. Л. Чебышев, А. А. Марков, А. М. Ляпунов). Широкое применение вероятностных представлений в физике и в самых различных областях практической деятельности человека привело к тому, что к началу XX века назрела необходимость уточнения понятия вероятности. Это было необходимо, в частности, для того, чтобы избежать спекуляций, связанных с неоправданными применениями понятия вероятности, опирающимися лишь на «житейские» представления. Неудачную попытку дать общее определение вероятности случайного события на основе предела частоты его появления предпринял немецкий математик Р. Мизес. Уточнение понятия вероятности произошло на основе не частотного, а *аксиоматического* подхода. Такой подход основывается на некоторых первичных положениях (аксиомах), из которых выводятся все остальные положения в результате применения определенных четко сформулированных правил.

Общепринятое сегодня *аксиоматическое* определение вероятности было разработано советским математиком академиком А. Н. Колмогоровым и изложено им в книге «Основные понятия теории вероятностей» (1936 г.). Мы не будем рассматривать аксиоматическое определение вероятности, так как для этого пришлось бы обратиться к теории множеств. Отметим лишь, что предложенная А. Н. Колмогоровым аксиоматика поставила понятие вероятности на строгую математическую основу, в результате чего теория вероятностей окончательно укрепилась как полноправная математическая дисциплина.

Существование нескольких определений для одного и того же понятия (вероятности) не должно удивлять читателя.

«Многообразие определений основных понятий — существенная черта современной науки, и понятие вероятности не исключение,— пишет Л. Е. Майстров в книге «Развитие понятия вероятности» (М.: Наука, 1980).— Современные определения в науке — это изложение точек зрения, которых может быть много для любого фундаментального понятия, и все они отражают какую-нибудь существенную сторону определяемого понятия. Это относится и к понятию вероятности». Добавим, что появление новых определений данного понятия происходит по мере углубления наших представлений об этом понятии, выявления новых его сторон.

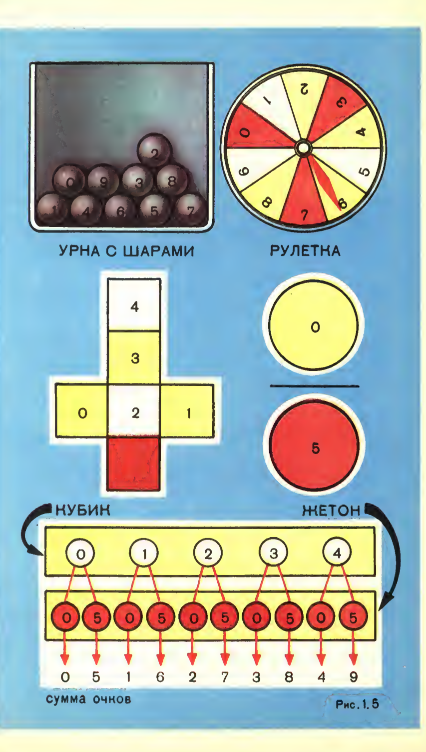
Случайные числа

Генераторы случайных чисел. Положим в ящик десять одинаковых шаров, помеченных цифрами от 0 до 9. Вынем наугад один из шаров и отметим его цифру. Пусть это будет 5. Затем вернем шар в ящик, хорошо перемешаем шары и снова вынем наугад один шар. Допустим, что на этот раз выпала цифра 1. Запишем ее, вернем шар в ящик, перемешаем шары, и снова вынем наугад один шар. Предположим, что выпала цифра 2. Повторяя эту операцию много раз, мы получим неупорядоченный набор цифр, например такой: 5, 1, 2, 7, 2, 3, 0, 2, 1, 3, 9, 2, 4, 4, 1, 3... Неупорядоченность набора связана с тем, что каждая цифра выпадала *случайно*. Ведь всякий раз шар вынимали наугад из хорошо перемешанной совокупности одинаковых шаров.

Имея набор случайных цифр, можно составить набор *случайных чисел*. Будем рассматривать, например, четырехзначные числа. В этом случае достаточно разбить полученный набор случайных цифр на группы по четыре цифры и рассматривать каждую группу как одно из таких чисел: 5127, 2302, 1392, 4413...

Устройства для получения наборов случайных чисел называют *генераторами случайных чисел*. Различают три типа таких генераторов: *урны, кости, рулетки*. Рассмотренный только что ящик с шарами представляет собой одну из разновидностей урн. Другая разновидность — лототрон, используемый в телепередачах спортлото.

Наиболее просто устроены генераторы случайных чисел, относящиеся к типу «кости». Примерами таких генераторов являются подбрасываемый кубик, грани которого помечены разными цифрами, подбрасываемая монета (или жетон) и т. д. Предположим, что пять граней кубика помечены цифрами 0, 1, 2, 3, 4, а шестая закрашена. Предположим также, что имеется жетон, на одной стороне которого проставлена цифра 0, а на другой 5. Будем подбрасывать кубик и жетон одновременно и всякий раз подсчитывать сумму выпавших очков. (Если кубик падает закрашенной гранью, то такое испытание не принимается во внимание.) Такой генератор позволяет получить неупорядоченный набор цифр, включающий все цифры от 0 до 9, из которого нетрудно затем образовать набор (наборы) случайных чисел.



2
0
9
3
8
1
4
6
5
7
УРНА С ШАРАМИ
1
2
3
4
5
6
7
8
9
0
РУЛЕТКА
4
3
0
2
1
0
5
КУБИК
ЖЕТОН
0
1
2
3
4
0
5
0
5
0
5
0
5
0
5
0
5
1
6
2
7
3
8
4
9
сумма очков

Рис. 1.5

Еще один тип генераторов случайных чисел — рулетка. Она представляет собой круг, разбитый на некоторое число секторов, каждому из которых соответствует определенная цифра (или число). Рулетка имеет вращающуюся по кругу стрелку (или катящийся шарик). Испытание состоит в том, чтобы привести в движение (толкнуть) стрелку и подождать, когда она остановится. Результатом испытания является цифра, соответствующая тому сектору рулеточного круга, в пределах которого остановилась стрелка. Рулетка с вращающейся стрелкой используется, например, в телепередаче «Что? Где? Когда?».

Заметим, что число секторов, на которые разбивается рулеточный круг, может быть каким угодно. В частности, можно разбить круг на 10 одинаковых секторов и пометить их цифрами от 0 до 9. В этом случае рулетка как генератор случайных чисел будет эквивалентна двум рассмотренным ранее генераторам — урне с десятью шарами и подбрасываемым одновременно кубику и жетону. На рисунке 1.5 схематически изображены все эти эквивалентные друг другу генераторы случайных чисел.

Таблица случайных чисел. Пример таблицы случайных чисел приведен на рисунке 1.6. Таблица состоит из трехсот четырехзначных чисел. Каждая цифра в таблице появилась *случайным образом* — как результат некоего испытания, например подбрасывания кубика и жетона. Поэтому понятна неупорядоченность появления цифр, не позволяющая предсказать цифру, которая последует за только что выпавшей. Выполнив соответствующее число испытаний, вы можете составить много подобных таблиц. И все равно вам не удастся обнаружить и тени какого-либо порядка в следовании цифр друг за другом.

Все это неудивительно. Ведь на то он и случай! Но у случая есть и оборотная сторона. Посчитайте, например, сколько раз в таблице на рисунке 1.6 встречается та или иная цифра. Вы обнаружите, что цифра 0 встречается 118 раз (частота появления равна 118/1200=0,099), цифра 1 встречается 110 раз (частота появления 0,090), цифра 2 — 114 раз (0,095), цифра 3—125 раз (0,104), цифра 4—135 раз (0,113), цифра 5—135 раз (0,113), цифра 6—132 раза (0,110), цифра 7—116 раз (0,097), цифра 8—93 раза (0,078), цифра 9—122 раза (0,102). Можно видеть, что частота появления каждой цифры примерно одна и та же: она близка к 0,1. Читатель, конечно, сообразил, что 0,1 — это *вероятность* выпадения той или иной конкретной цифры. Естественно, может сказать он, что при достаточно большом числе испытаний (а их здесь 1200) частота появления той или иной цифры оказывается близкой к вероятности ее выпадения.

И хотя все это естественно, нельзя не подивиться лишний раз тому, как в неупорядоченном наборе *случайно появляющихся* цифр обнаруживается *внутренняя устойчивость*. Здесь наглядно проявляется оборотная сторона случая, принимающая облик точно определяемой *вероятности*.

Посоветуем читателю немного «поработать» с таблицей случай-

0655	8453	4467	3384	5320	0709	2523	9224	6271	2607
5255	5161	4889	7429	4647	4331	0010	8144	8638	0307
6314	8951	2335	0174	6993	6157	0063	6006	1736	3775
3157	9764	4862	5848	6919	3135	2837	9910	7791	8941
9052	9565	4635	0653	2254	5704	8865	2627	7959	3682
4105	4105	3187	4312	1596	9403	6859	7802	3180	4499
1437	2851	6727	5580	0368	4746	0604	7956	2304	8417
4064	4171	7013	4631	8288	4785	6560	8851	9928	2439
1037	5765	1562	9869	0756	5761	6346	5392	2986	2018
5718	8791	0754	2222	2013	0830	0927	0466	7526	6610
5127	2302	1392	4413	9651	8922	1023	6265	7877	4733
9401	2423	6301	2611	0650	0400	5998	1863	9182	9032
4064	5228	4153	2544	4125	9854	6380	6650	8567	5045
5458	1402	9849	9886	5579	4171	9844	0159	2260	1314
2461	3497	9785	5678	4471	2873	3724	8900	7852	5843
4320	4558	2545	4436	9265	6675	7989	5592	3759	3431
3466	8269	9926	7429	7516	1126	6345	4576	5059	7746
9313	7489	2464	2575	9284	1787	2391	4245	5618	0146
5179	8081	3361	0109	7730	6256	1303	6503	4081	4754
3010	5081	3300	9979	1970	6279	6307	7935	4977	0501
9599	9828	8740	6666	6692	5590	2455	3963	6463	1609
4242	3961	6247	4911	7264	0247	0583	7679	7942	2482
3585	9123	5014	6328	9659	1863	0532	6313	3199	7619
5950	3384	0276	4503	3333	8967	3382	3016	0639	2007
8462	3145	6582	8605	7300	6298	6673	6406	5951	7427
0456	0944	3058	2545	3756	2436	2408	4477	5707	5441
0672	1281	8697	5409	0653	5519	9720	0111	4745	7979
5163	9690	0413	3043	1014	0228	5460	2835	3294	3674
4995	9115	5273	1293	7894	9050	1378	2220	3756	9795
6751	6447	4991	6458	9307	3371	3243	2958	4738	3996

Рис.1.6

ных чисел (см. рис. 1.6). Он может убедиться, например, что из трехсот приведенных в таблице чисел 32 начинаются с нуля, 20 — с единицы, 33 — с двойки, 33 — с тройки, 38 — с четверки, 34 — с пятерки, 34 — с шестерки, 24 — с семерки, 20 — с восьмерки, 32 — с девятки. Вероятность того, что число будет начинаться с той или иной конкретной цифры, равна 0,1. Легко видеть, что приведенные результаты неплохо согласуются с этой вероятностью (одна десятая от трехсот есть тридцать). Правда, отклонения оказываются более значительными, чем в примере, который рассматривался ранее. Но это естественно, поскольку в том примере число испытаний равнялось 1200, тогда как здесь оно заметно меньше: всего лишь 300.

Интересно также подсчитать, сколько раз та или иная цифра встречается на втором месте (количество сотен в числе), на третьем месте (десятки), на четвертом месте (единицы). Нетрудно убедиться, что во всех случаях частота появления выбранной цифры близка к вероятности, т. е. близка к 0,1. Так, нуль встречается на втором месте 25 раз, на третьем — 33 раза, на четвертом — 28 раз.

В вводной беседе приводился пример с номерами автомобилей, случайно проезжающих мимо наблюдателя. Как отмечалось, вероятность того, что первые две цифры номера окажутся одинаковыми, равна 0,1. Такова же вероятность оказаться одинаковыми двум последним цифрам номера, или двум крайним цифрам, или, наконец, двум средним цифрам.

Чтобы убедиться в этом, совсем не обязательно дежурить на обочине шоссе. Достаточно обратиться к таблице случайных чисел (см. рис. 1.6). Четырехзначные случайные числа в таблице могут рассматриваться как номера случайно проехавших мимо наблюдателя автомобилей. Мы видим, что из трехсот номеров имеют одинаковые две первые цифры 40 номеров, одинаковые две последние цифры 28 номеров, одинаковые крайние цифры 32 номера, одинаковые средние цифры 24 номера. Иначе говоря, частоты появления пар одинаковых цифр действительно колеблются вблизи вероятности, т. е. вблизи 0,1.

Случайные события

Когда мы фиксируем выпадение того или иного количества очков при подбрасывании кубика или при вытаскивании шара из урны, мы всякий раз имеем дело со *случайными событиями*. Рассмотрим несколько занимательных задач, где требуется вычислить вероятность случайного события.

Задача с разноцветными шарами. В ящике находятся три синих и один красный шар. Вы наугад вынимаете из ящика два шара. Какая вероятность больше — вынуть два синих шара или вынуть синий и красный шары?

На этот вопрос часто отвечают, что более вероятно вынуть два синих шара, поскольку синих шаров в ящике в три раза больше,

чем красных. В действительности же вероятность вынуть два синих шара *равна* вероятности вынуть синий и красный шары. В этом можно убедиться, посмотрев на рисунок 1.7. Из рисунка видно, что существуют три способа вынуть два синих шара и три способа вынуть синий и красный шары. Следовательно, рассматриваемые исходы равновероятны.

Наконец, можно вычислить вероятность исходов. Вероятность вынуть два синих шара равна произведению двух вероятностей. Первая есть вероятность вынуть синий шар из совокупности четырех шаров (три синих плюс один красный); она равна 3/4. Вторая есть вероятность вынуть синий шар из совокупности трех шаров (два синих плюс один красный); она равна 2/3. Таким образом, вероятность вынуть подряд два синих шара равна $3/4 \times 2/3 = 1/2$.

Вероятность вынуть синий и красный шары может быть представлена в виде суммы $P_{ск} + P_{кс}$, где $P_{ск}$ — вероятность вынуть синий шар из совокупности четырех шаров (три синих плюс один красный), умноженная на вероятность вынуть красный шар из совокупности трех шаров (два синих плюс один красный), а $P_{кс}$ — вероятность вынуть красный шар из совокупности четырех шаров (в этом случае второй шар будет с достоверностью синим). Иначе говоря, $P_{ск}$ — вероятность сначала вынуть синий, а затем красный шар, тогда как $P_{кс}$ — вероятность сначала вынуть красный, а затем синий шар. Поскольку $P_{ск} = 3/4 \times 1/3 = 1/4$ и $P_{кс} = 1/4$, то, следовательно, вероятность вынуть пару разноцветных шаров равна $1/4 + 1/4 = 1/2$.

Игра с бросанием кубика. В этой игре участвуют два игрока — игрок *А* и игрок *Б*. За один ход кубик бросают три раза подряд. Если при этом хотя бы один раз выпадет определенная грань (пусть это будет грань с единицей), то игрок *А* записывает себе очко. Если же единица не выпадет ни разу, то очко записывается игроку *Б*. Поочередно игроки выполняют по три бросания до тех пор, пока один из них не наберет, скажем, сто очков. У кого больше шансов выиграть — у игрока *А* или у игрока *Б*?

Чтобы ответить на этот вопрос, вычислим вероятность того, что игрок *А* получит очко в результате трех бросаний кубика. Он получает очко в любом из следующих трех случаев: если единица выпала в первом же бросании; если единица не выпала в первом бросании, но выпала во втором; если единица не выпала в первых двух бросаниях, но выпала в третьем. Обозначим вероятности реализации этих трех случаев соответственно через P_1, P_2, P_3. Искомая вероятность есть $P = P_1 + P_2 + P_3$. Заметим, что вероятность выпадения единицы в каждом бросании равна 1/6, а вероятность того, что единица не выпадет, равна 5/6. Ясно, что $P_1 = 1/6$. Для нахождения P_2 надо умножить вероятность невыпадения единицы (в первом бросании) на вероятность ее выпадения (во втором бросании): $P_2 = 5/6 \times 1/6 = 5/36$. Вероятность P_3 равна произведению вероятности невыпадения единицы в двух бросаниях (в первом и втором) и вероятности выпадения единицы в одном

бросании (в третьем): $P_3=(5/6)^2 \times 1/6=25/216$. Таким образом, $P=P_1+P_2+P_3=1/6+5/36+25/216=91/216$. Поскольку $P<1/2$, то, следовательно, в данной игре более вероятен выигрыш игрока *Б*. К такому же выводу можно прийти проще, если рассматривать вероятность того, что игрок *Б* получит очко в результате трех бросаний кубика. Это есть вероятность невыпадения единицы в трех бросаниях: $p=5/6 \times 5/6 \times 5/6=125/216$. Так как $p>1/2$, то шансы игрока *Б* оказываются более предпочтительными. Заметим, что $P+p=91/216+125/216=1$. Это естественно, так как за каждый ход кто-нибудь обязательно получает очко — либо один игрок, либо другой.

Немного изменим правила игры: пусть за один ход выполняются не *три*, а *четыре* бросания кубика. Остальные условия остаются прежними. В этом случае вероятность того, что игрок *Б* получит очко за один ход, равна $5/6 \times 5/6 \times 5/6 \times 5/6=625/1296$. Она меньше 1/2, следовательно, теперь больше шансов выиграть уже не у игрока *Б*, а у игрока *А*.

Задача о звездочете. Некий властелин разгневался на звездочета и повелел палачу отрубить ему голову. Однако в последний момент властелин смягчился и решил дать звездочету возможность спастись. Он взял два черных и два белых шара и предложил звездочету произвольным образом распределить их по двум урнам. Палач должен выбрать наугад одну из урн и наугад вытащить из нее шар. Если шар окажется белым, то звездочет будет помилован, а если черным, казнен. Как должен звездочет распределить шары по двум урнам, чтобы иметь наибольшее число шансов спастись?

Допустим, что звездочет положит в каждую урну по одному белому и одному черному шару (рис. 1.8, *а*). В этом случае безразлично, к какой урне подойдет палач. Из любой урны он с вероятностью 1/2 вынет белый шар. Значит, вероятность спастись звездочету равна 1/2.

Такова же будет вероятность спастись, если звездочет положит в одну урну два белых шара, а в другую два черных (рис. 1.8, *б*). Все решит выбор палачом той или иной урны. Палач с равной вероятностью может подойти как к «белой», так и к «черной» урне.

Лучше всего, если звездочет положит в одну урну белый шар, а в другую белый и два черных (рис. 1.8, *в*). Если палач подойдет к первой урне, то звездочет спасется наверняка. Если же палач подойдет ко второй урне, то звездочет будет иметь вероятность спастись, равную 1/3. Так как вероятность того, что палач подойдет к той или иной урне, равна 1/2, то полная вероятность звездочету спастись может быть вычислена следующим образом: $(1/2 \times 1)+(1/2 \times 1,3)=2/3$.

Если же звездочет положит в одну урну черный шар, а в другую черный и два белых (рис. 1.8, *г*), то вероятность спастись окажется наименьшей: $(1/2 \times 0)+(1/2 \times 2/3)=1/3$.

Итак, чтобы иметь наибольшие шансы спастись, звездочет должен

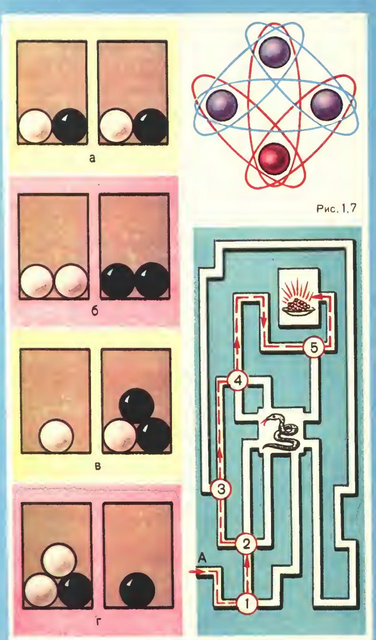


Рис. 1.7

Рис. 1.8

Рис. 1.9

избрать вариаит распределения шаров по урнам, показанный иа рисуике 1.8, *в*. Это есть наилучшая тактика. Наихудшая тактика отвечает вариаиту распределения шаров, показаниому иа рисуике 1.8, *г*. Разумеется, выбор наилучшей тактики не гарантирует спасения. Риск хотя и уменьшается, но все же остается.

Блуждание в лабиринте. На рисунке 1.9 изображеи лабиринт, в котором хранятся сокровища и имеется западия. Неудачливые охотиики за сокровищами, попадая в западню, погибают. Какова вероятиость избежать западни и добраться до сокровищ?

Пройдя от входа *A* до пункта *1* (см. рис. 1.9), искатель сокровищ может пойти прямо (тогда он сразу же попадает в западню) или повериуть иалево (тогда он попадает в пуикт *2*). Будем полагать, что выбор того или иного из этих двух вариантов осуществляется с одиой и той же вероятиостью, т. е. с вероятиостью 1/2. Попав в пуикт *2*, искатель сокровищ с вероятиостью 1/3 выбирает далее путь либо прямо, либо направо, либо иалево. Первые два пути приводят в западню, а третий приводит в пуикт *3*. Вероятиость попасть от входа *A* в пункт *3* равна произведеиию вероятиости повериуть в пуикте *1* иалево и вероятиости повериуть в пуикте *2* также иалево: $1/2\times1/3$. Нетрудно далее сообразить, что вероятиость добраться от *A* до пункта *4* равна $1/2\times1/3\times1/2$; вероятиость добраться от *A* до пункта *5* равиа $1/2\times1/3\times1/2\times1/3$; иакоиец, вероятиость попасть из *A* в хранилище сокровищ равиа $P^{+}=1/2\times1/3\times1/2\times1/3\times1/2=1/72$. Единствениый путь виутри лабиринта от входа до сокровищ показаи иа рисунке штриховой лиииеи. Ои реализуется с вероятиостью $P^{+}=1/72$. С вероятиостью $P^{-}=71/72$ искатель сокровищ попадает в западню.

Вероятиость P^{-} была определена исходя из того, что $P^{+}+P^{-}=1$. Можно вычислить P^{-} непосредственно. Представим P^{-} в виде суммы $P^{-}=P_1+P_2+P_3+P_4+P_5$, где P_i есть произведение вероятиости попасть из *A* в пуикт *i* и вероятности попасть из пункта *i* в западию ($i=1, 2, 3, 4, 5$).

$P_1=1/2$,
$P_2=1/2\times2/3$,
$P_3=1/2\times1/3\times1/2$,
$P_4=1/2\times1/3\times1/2\times2/3$,
$P_5=1/2\times1/3\times1/2\times1/3\times1/2$.
Нетрудио убедиться, что $P_1+P_2+P_3+P_4+P_5=71/72$

Дискретные случайные величины

Случайные величиины. Предположим, что в иекоторой партии из 100 изделий забраковаио 11 изделий; в другой такой же партии забраковано 9 изделий, в третьей 10 изделий, в четвертой 12 изделий. И так далее. Обозиачим через *n* полное число изделий в партии, а через *m* число бракованиых изделий. Величина *n* постояииа (здесь $n=100$), а величина *m* изменяется от партии к партии случайным образом. Будем полагать, что существует

определенная вероятность появления m бракованных изделий в наугад выбранной партии из n изделий.

Количество бракованных изделий (величина m) является примером *случайной величины*. Это есть величина, *значения которой изменяются случайным образом от одного испытания к другому, причем каждое из значений реализуется с той или иной вероятностью*. Подчеркнем, что речь идет о *дискретной* случайной величине, т. е. величине, возможные значения которой образуют дискретный набор чисел (в данном случае целочисленные значения от 1 до 100).

Существуют также *непрерывные* случайные величины. Например, рост и вес новорожденных изменяются случайно от одного ребенка к другому, принимая любые значения в некотором интервале. Рассмотрение непрерывных случайных величин имеет свои особенности. Мы остановимся на них позднее, а пока ограничимся дискретными величинами.

Математическое ожидание и дисперсия дискретной случайной величины. Пусть x — некоторая дискретная случайная величина, которая может принимать s значений: $x_1, x_2, ..., x_m, ..., x_s$. Этим значениям соответствуют вероятности: $p_1, p_2, ..., p_m, ..., p_s$. Например, p_m есть вероятность того, что рассматриваемая величина примет x_m. Сумма всех вероятностей $(p_1+p_2+...+p_s)$ есть вероятность того, что в испытании будет реализовано какое-либо (безразлично, какое именно) из значений $x_1, x_2, ..., x_s$. Эта вероятность равна единице. Таким образом,

$$\sum_{m=1}^{s} p_m = 1 \qquad (1.3)$$

(знак $\sum_{m=1}^{s}$ означает, что выполняется суммирование по всем m от 1 до s).

Набор вероятностей $p_1, p_2, ..., p_s$ (говорят также о распределении вероятностей) содержит исчерпывающую информацию о случайной величине. Однако во многих случаях на практике знание вероятностей необязательно. Достаточно знать две наиболее важные характеристики случайной величины — ее математическое ожидание и дисперсию.

Математическое ожидание есть среднее значение случайной величины. Усреднение производится по большому числу испытаний. Для обозначения таких средних будем использовать скобки $\langle ... \rangle$. Среднее случайной величины x есть сумма произведений значений этой величины на соответствующие вероятности:

$$\langle x \rangle = p_1x_1 + p_2x_2 + ... + p_sx_s,$$

или, если использовать знак суммирования,

$$\langle x \rangle = \sum_{m=1}^{s} p_m x_m. \qquad (1.4)$$

Кроме среднего значения, важно также знать, насколько сильно значения рассматриваемой величины отклоняются от ее среднего значения или, иначе говоря, насколько широк *разорос* значений случайной величины. Среднее значение отклонения от среднего (среднее значение разности $x-\langle x\rangle$) здесь не годится, поскольку оно равно нулю. Действительно,

$$\langle(x-\langle x\rangle)\rangle=\sum_{m=1}^{s}p_m(x_m-\langle x\rangle)=\sum_{m=1}^{s}p_mx_m-\langle x\rangle\sum_{m=1}^{s}p_m=$$
$$=\langle x\rangle-\langle x\rangle=0.$$

Поэтому рассматривают среднее значение не самого отклонения от среднего, а *квадрата* отклонения, т. е. рассматривают

$$D=\langle(x-\langle x\rangle)^2\rangle=\sum_{m=1}^{s}p_m(x_m-\langle x\rangle)^2. \quad (1.5)$$

Это и есть *дисперсия* случайной величины, будем обозначать ее через D. Квадратный корень из дисперсии ($\sqrt{D}$) называют *средним квадратичным отклонением* или *стандартом* случайной величины. Нетрудно убедиться, что

$$D=\langle x^2\rangle-\langle x\rangle^2. \quad (1.6)$$

Действительно,

$$\sum_{m=1}^{s}p_m(x_m-\langle x\rangle)^2=\sum_{m=1}^{s}p_m(x_m^2-2x_m\langle x\rangle+\langle x\rangle^2)=$$
$$=\sum_{m=1}^{s}p_mx_m^2-2\langle x\rangle\sum_{m=1}^{s}p_mx_m+\langle x\rangle^2\sum_{m=1}^{s}p_m=$$
$$=\langle x^2\rangle-2\langle x\rangle\langle x\rangle+\langle x\rangle^2=\langle x^2\rangle-\langle x\rangle^2.$$

На рисунке 1.10, *а* сопоставляются два распределения вероятностей. Сопоставляемые случайные величины имеют различное математическое ожидание, но одинаковую дисперсию. На рисунке 1.10, *б*, наоборот, сопоставляемые случайные величины имеют различную дисперсию, но одинаковое математическое ожидание.

Формула Бернулли (биномиальное распределение). Допустим, что выполняется серия из n независимых одинаковых испытаний. Испытания независимы в том смысле, что результаты одних испытаний не влияют на результаты других. Некоторые из них дают интересующий нас исход, остальные не дают. Будем называть интересующий нас исход событием U. Это — случайное событие. Пусть m — число испытаний, в которых реализовалось событие U. Это — случайная величина. Будем рассматривать вероятность $P_n(m)$ того, что в серии из n испытаний событие U реализуется m раз.

Такая схема описывает многие реальные ситуации: проверяются n изделий, событие U — появление изделия с браком, $P_n(m)$ — вероятность того, что среди n изделий окажется m бракованных. Реги-

стрируются n новорожденных, событие U — рождение девочки, $P_n(m)$ — вероятность того, что среди n новорожденных будет m девочек. Проверяются n лотерейных билетов, событие U — выигрыш, $P_n(m)$ — вероятность того, что из n билетов выиграют m. В физическом эксперименте регистрируются n нейтронов, событие U — регистрация нейтрона с энергией в некотором интервале значений, $P_n(m)$ — вероятность того, что из n нейтронов будут иметь энергию в рассматриваемом интервале значений m нейтронов. Во всех этих случаях вероятность $P_n(m)$ описывается одной и той же формулой — *формулой Бернулли* (по имени швейцарского математика XVII века Якоба Бернулли).

В основе вывода формулы Бернулли лежит предположение, что вероятность появления события U в единичном испытании известна и не меняется от испытания к испытанию. Обозначим эту вероятность через p. Тогда вероятность непоявления события U в единичном испытании есть $q=1-p$. Обратим внимание на то, что вероятность появления бракованного изделия нисколько не зависит от того, сколько бракованных изделий в данной партии уже появилось. Вероятность рождения девочки в том или ином конкретном случае не зависит от того, кто родился перед этим — девочка или мальчик (а также от того, сколько девочек уже родилось). Вероятность выигрыша не увеличивается и не уменьшается по мере проверки лотерейных билетов. Вероятность появления нейтрона с энергией в заданном интервале значений не меняется в течение эксперимента.

Итак, *знная вероятность p появления некоторого случайного события в единичном испытании, найдем вероятность $P_n(m)$ того, что в серии из n независимых одинаковых испытаний это событие появится m раз.*

Допустим, что интересующее нас событие U появилось в первых m испытаниях и не появилось в оставшихся $n-m$ испытаниях. Вероятность такой ситуации равна $p^m q^{n-m}$. Разумеется, возможен иной порядок появления события U. Например, оно может не появиться в первых $n-m$ испытаниях и появиться в оставшихся m испытаниях. Вероятность такого варианта также равна $p^m q^{n-m}$. Возможны и другие варианты. Всего таких вариантов столько, сколько имеется *сочетаний* из n элементов по m (это число сочетаний обозначают через C_n^m). Каждый из вариантов имеет вероятность $p^m q^{n-m}$. Для нас безразличен порядок появления события U. Важно, чтобы оно появилось в каких-то m испытаниях и не появилось в остальных $n-m$ испытаниях. Поэтому для нахождения искомой вероятности $P_n(m)$ надо сложить вероятности реализации всех C_n^m вариантов, т. е. умножить $p^m q^{n-m}$ на C_n^m:

$$P_n(m)=C_n^m p^m q^{n-m}. \quad (1.7)$$

Существует формула для числа сочетаний из n элементов по m:

$$C_n^m=\frac{n!}{m!(n-m)!}=\frac{n(n-1)(n-2)...(n-m+1)}{m!}. \quad (1.8)$$

Здесь $n!=1\cdot2\cdot3\cdot\ldots\cdot n$ (говорят: «эн факториал»), причем $0!=1$. Подставляя формулу (1.8) в (1.7), находим

$$P_n(m)=\frac{n!}{m!(n-m)!}\,p^m q^{n-m}. \tag{1.9}$$

Это и есть *формула Бернулли.* Ее называют также биномиальным распределением вероятностей случайных величин или *биномиальным распределением.* Ниже мы поясним происхождение такого названия и заодно убедимся, что

$$\sum_{m=0}^{n} P_n(m)=1. \tag{1.10}$$

В качестве примера рассмотрим вероятность рождения m девочек в группе из 20 новорожденных. Вероятность рождения девочки в «единичном испытании» примем равной 1/2. В данном случае в выражении (1.9) надо принять $p=1/2$, $n=20$ и рассматривать целочисленные значения величины m в интервале от 0 до 20. Результат удобно представить графически (рис. 1.11). Мы видим, что наиболее вероятно рождение 10 девочек. Вероятность рождения, например, 6 или 14 девочек уже в шесть раз меньше. Если случайная величина описывается биномиальным распределением, то ее математическое ожидание $\left[\langle m\rangle=\sum_{m=0}^{n} mP_n(m)\right]$ равно произведению числа испытаний и вероятности появления события в единичном испытании,

$$\langle m\rangle=np. \tag{1.11}$$

Дисперсия такой случайной величины равна произведению трех сомножителей — числа испытаний, вероятности появления события в единичном испытании и вероятности непоявления события:

$$D=\langle m^2\rangle-\langle m\rangle^2=npq. \tag{1.12}$$

Нормальное распределение (распределение Лапласа — Гаусса). При больших n использование формулы Бернулли связано с трудоемкими вычислениями. Чтобы, например, найти вероятность того, что из 50 новорожденных будет 30 девочек, надо вычислить

$$P_{30}(50)=\frac{50!}{30!20!}(0{,}5)^{50}.$$

Заметим, что уже 20! представляет собой число из 19 цифр!
В подобных случаях можно пользоваться формулой, которая представляет собой *предельный* случай формулы Бернулли при достаточно больших n. Эта формула имеет вид:

$$P_n(m)=\frac{1}{\sqrt{2\pi D}}\,e^{-(m-\langle m\rangle)^2/2D}, \tag{1.13}$$

где $\langle m\rangle=np$, $D=npq$. Используемое здесь число $e=2{,}718\ldots$ есть

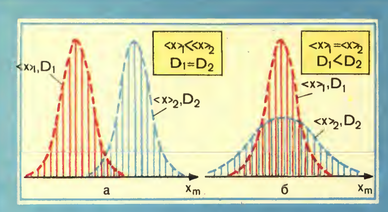


Рис. 1.10

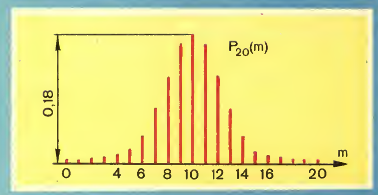


Рис. 1.11

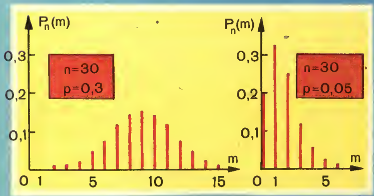


Рис. 1.12

основание натуральных логарифмов. Распределение (1.13) называют *нормальным распределением* или *распределением Лапласа — Гаусса.*

Распределение Пуассона. Если вероятность появления в единичном испытании интересующего нас события *очень мала* ($p \ll 1$), то при больших n биномиальное распределение переходит не в нормальное распределение, а в *распределение Пуассона:*

$$P_n(m) = \frac{(np)^m}{m!}\, e^{-np}. \qquad (1.14)$$

Это распределение называют также *законом редких явлений.* Полезно заметить, что дисперсия случайной величины, описываемой распределением Пуассона, равна ее математическому ожиданию.

На рисунке 1.12 сопоставляются два распределения $P_n(m)$. Первое имеет параметры $n=30$, $p=0{,}3$; оно близко к нормальному распределению с математическим ожиданием $\langle m \rangle = 9$. Второе имеет параметры $n=30$, $p=0{,}05$; оно близко к распределению Пуассона с $\langle m \rangle = 1{,}5$.

Немного математики. Прежде всего остановимся на происхождении названия «биномиальное распределение». Алгебраическое выражение $(q+p)^n$, где n — целое положительное число, называют биномом Ньютона n-й степени. Читателям хорошо известны биномы Ньютона второй и третьей степени:

$$(q+p)^2 = q^2 + 2qp + p^2;$$

$$(q+p)^3 = q^3 + 3q^2p + 3qp^2 + p^3 .$$

В общем случае (при произвольном целочисленном n) бином Ньютона описывается формулой:

$$(q+p)^n = q^n + nq^{n-1}p + \ldots + \frac{n(n-1)\ldots(n-m+1)}{m!}\, q^{n-m}p^m + \ldots + nqp^{n-1} + p^n.$$

С учетом (1.8) можно переписать эту формулу в виде

$$(q+p)^n = C_n^0 q^n + C_n^1 q^{n-1}p + \ldots + C_n^m q^{n-m}p^m + \ldots + C_n^{n-1} qp^{n-1} + C_n^n p^n.$$

В соответствии с (1.9) заключаем, что

$$(q+p)^n = \sum_{m=0}^{n} C_n^m q^{n-m} p^m = \sum_{m=0}^{n} P_n(m).$$

Таким образом, вероятности $P_n(m)$ совпадают с членами разложения бинома Ньютона. Отсюда и происходит название — *биномиальное распределение.*

Входящие в биномиальное распределение вероятности q и p удовлетворяют условию $q+p=1$. Следовательно, $(q+p)^n = 1$. С другой стороны, $(q+p)^n = \sum_{m=0}^{n} P_n(m)$. Отсюда приходим к результату (1.10).

Непрерывные случайные величины

Рассмотрение непрерывных случайных величин имеет свои особенности. Непрерывная величина принимает бесконечное множество значений, которые сплошь заполняют некоторый промежуток. Принципиально невозможно перечислить все значения такой величины хотя бы уже потому, что нельзя указать два соседних значения (подобно тому как нельзя указать на числовой оси две соседние точки). Кроме того, вероятность каждого конкретного значения непрерывной случайной величины равна нулю.

Может ли равняться нулю вероятность возможного события? Читатель знает, что равную нулю вероятность имеют невозможные события. Оказывается, что и возможные события могут иметь нулевую вероятность.

Предположим, что на полоску бумаги, на которой изображен отрезок числовой оси, произвольно бросают много раз тонкую иглу. В качестве непрерывной случайной величины может рассматриваться координата x той точки, в которой игла пересекает числовую ось (рис. 1.13, *а*). Эта координата меняется от одного бросания иглы к другому случайным образом.

Вместо того чтобы бросать иглу, можно воспользоваться рулеткой. На окружность рулеточного круга наклеим полоску бумаги с отрезком числовой оси, как это показано на рисунке 1.13, *б* (будем полагать, что длина отрезка равна длине окружности). Различные значения непрерывной случайной величины x будут реализовываться в данном случае не в результате бросания иглы, а при остановке свободно вращающейся стрелки рулетки.

Чему равна вероятность того, что стрелка остановится напротив некоторой конкретной точки x? Иными словами, чему равна вероятность реализации конкретного значения x непрерывной случайной величины? Предположим, что рулеточный круг радиуса R разбит на конечное число одинаковых секторов, например на 10 секторов (рис. 1.14). Длина дуги сектора равна $\Delta x = 2\pi R/10$. Вероятность того, что стрелка остановится в пределах заштрихованного на рисунке сектора, равна $\Delta x/2\pi R = 1/10$. Итак, вероятность реализации какого-нибудь значения случайной величины в промежутке от x до $x + \Delta x$ равна $\Delta x/2\pi R$. Будем постепенно уменьшать Δx, т. е. будем разбивать круг на все большее и большее число секторов. Соответственно будет уменьшаться вероятность $\Delta x/2\pi R$ реализации какого-нибудь значения из промежутка от x до $x + \Delta x$. Чтобы получить вероятность реализации значения, равного x, надо перейти к пределу при $\Delta x \to 0$. В этом случае вероятность $\Delta x/2\pi R$ обращается в нуль. Мы убеждаемся, что вероятность реализации того или иного конкретного значения непрерывной случайной величины действительно равна нулю.

Представление о событии, являющемся возможным и в то же время обладающем нулевой вероятностью, может показаться парадоксальным. В действительности же никакого парадокса нет. С подобными ситуациями читатель наверняка знаком. Рассмотрим одну из

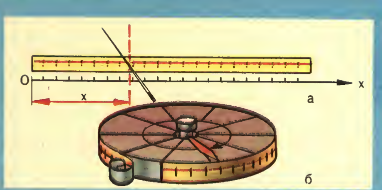


Рис. 1. 13

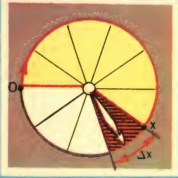


Рис. 1. 14

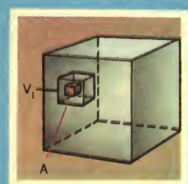


Рис. 1. 15

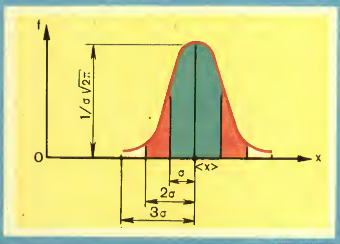


Рис. 1. 16

них в качестве примера. Пусть тело объемом V имеет массу M. Выделим внутри тела точку A и рассмотрим некоторый объем V_1, включающий эту точку (рис. 1.15). Пусть этому объему соответствует масса M_1. Будем постепенно уменьшать выделяемые внутри тела объемы, следя за тем, чтобы точка A все время оставалась внутри них. Мы получим последовательность объемов, стягивающуюся к точке A: V, V_1, V_2, V_3..., и соответствующую последовательность уменьшающихся масс: M, M_1, M_2, M_3... В пределе при стягивании объема к точке A масса обратится в нуль. Мы видим, что тело определенной массы состоит из точек с нулевой массой. Иными словами, ненулевая масса тела есть *сумма бесконечного числа нулевых масс* его отдельных точек. Точно так же ненулевая вероятность остановки стрелки рулетки в пределах промежутка Δx есть *сумма бесконечного числа нулевых вероятностей* остановки напротив отдельных значений x из рассматриваемого промежутка.

Плотность вероятности. Отмеченные выше затруднения можно обойти, используя понятие *плотности*. Масса отдельной точки тела равна нулю, но не равна нулю плотность в данной точке. Пусть ΔM — масса в объеме ΔV, внутри которого находится рассматриваемая точка (будем фиксировать ее радиус-вектором $\vec{r}$). Плотность $\varrho(\vec{r})$ в данной точке есть предел отношения $\Delta M / \Delta V$, получающийся при стягивании ΔV к точке $\vec{r}$:

$$\varrho(\vec{r}) = \lim_{\Delta V \to 0} \Delta M / \Delta V.$$

Если объем ΔV выбран достаточно малым, то можно приближенно считать, что $\Delta M \approx \varrho(\vec{r}) \Delta V$. При строгом подходе надо заменить ΔV дифференциалом dV. При этом масса M некоторого объема тела V будет выражаться *интегралом*:

$$M = \int_{(V)} \varrho(\vec{r}) dV,$$

который берется по всему рассматриваемому объему.

Аналогичным образом поступают и в теории вероятностей. Рассматривая *непрерывные* случайные величины, пользуются не понятием вероятности, а понятием *плотности вероятности*. Будем обозначать плотность вероятности случайной величины x через $f(x)$. По аналогии с обычной плотностью запишем

$$f(x) = \lim_{\Delta x \to 0} \Delta p_x / \Delta x.$$

Здесь через Δp_x обозначена вероятность реализации значений случайной величины в промежутке от x до $x + \Delta x$. Вероятность p реализации значений в промежутке от x_1 до x_2 выражается через плотность вероятности следующим образом:

$$p = \int_{x_1}^{x_2} f(x) dx. \qquad (1.15)$$

Если промежуток интегрирования охватывает все возможные значения случайной величины, то интеграл (1.15) равен единице (это есть вероятность достоверного события). В примере с рулеткой, который приводился выше, рассматриваемый промежуток есть промежуток от $x=0$ до $x=2\pi R$. В общем случае будем полагать этот промежуток бесконечным, так что

$$\int_{-\infty}^{\infty} f(x)dx=1. \qquad (1.16)$$

Заметим, что в примере с рулеткой рассматриваемый интеграл упрощается. Дело в том, что вероятность остановки стрелки рулетки в промежутке значений от x до $x+\Delta x$ *не зависит* от выбора x. Поэтому плотность вероятности не зависит в данном случае от x, следовательно,

$$\int_{0}^{2\pi R} f dx=f\int_{0}^{2\pi R} dx=2\pi R f=1; \qquad f=1/2\pi R.$$

С аналогичной ситуацией мы встречаемся, когда плотность тела одна и та же во всех его точках, т. е. когда тело *однородно* ($\varrho=M/V$). В общем же случае плотность $\varrho(\vec{r})$ меняется от одной точки тела к другой. Точно так же плотность вероятности $f(x)$ изменяется в общем случае при переходе от одного значения случайной величины к другому.

Математическое ожидание и дисперсия непрерывной случайной величины. Основные характеристики случайной величины — *математическое ожидание* и *дисперсия* — выражаются для дискретных величин через суммы по распределению вероятностей [см. формулы (1.4)—(1.6)]. Для непрерывных случайных величин вместо сумм используют интегралы, а вместо распределения вероятностей распределение плотности вероятности:

$$\langle x\rangle=\int_{-\infty}^{\infty} xf(x)dx; \qquad (1.17)$$

$$D=\int_{-\infty}^{\infty} (x-\langle x\rangle)^2 f(x)dx. \qquad (1.18)$$

Нормальное распределение плотности вероятности. Работая с непрерывными случайными величинами, мы часто встречаемся с *нормальным* распределением плотности вероятности. Это распределение описывается выражением [сравните с (1.13)]:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\,e^{-(x-\langle x\rangle)^2/2\sigma^2}. \qquad (1.19)$$

Здесь через σ обозначено среднее квадратичное отклонение ($\sigma=\sqrt{D}$). Функцию (1.19) называют *функцией (законом) Гаусса*.

Нормальным распределением описывается плотность вероятности

всех непрерывных случайных величин, разброс значений которых обусловлен множеством разнообразных факторов, действующих примерно в одинаковой степени и независимо друг от друга. В теории вероятностей доказывается, что сумма достаточно большого числа независимых случайных величин, подчиняющихся каким угодно законам распределения, приближенно описывается нормальным распределением, причем тем точнее, чем большее количество случайных величин суммируется.

Например, при массовом изготовлении гаек разброс значений их диаметра связан со случайными отклонениями характеристик материала, колебаниями температуры, вибрациями станка, изменениями напряжения в электросети, стачиванием инструмента и т. д. Все эти случайные факторы действуют примерно в одинаковой мере и независимо друг от друга. Они суммируются, и в результате диаметр гайки оказывается непрерывной случайной величиной, описываемой законом Гаусса. Математическое ожидание этой величины есть, очевидно, эталонное значение диаметра гайки, а дисперсия характеризует степень разброса реализуемых значений диаметра около эталонного значения.

«Правило трех сигм». Нормальное распределение плотности вероятности показано на рисунке 1.16. Максимум распределения достигается при значении x, равном математическому ожиданию $\langle x\rangle$. Кривая, описывающая рассматриваемое распределение (*кривая Гаусса*), имеет колоколообразный вид, она симметрична относительно вертикали $x=\langle x\rangle$. Площадь под кривой, рассматриваемая для всего бесконечного промежутка $(-\infty<x<+\infty)$, равна интегралу $\int\limits_{-\infty}^{\infty} f(x)\,dx$. Подставляя сюда функцию (1.19), можно убедиться, что эта площадь равна единице. Это согласуется с равенством (1.16), означающим, что вероятность достоверного события есть единица.

Разобьем площадь под кривой Гаусса вертикальными прямыми на отдельные участки (см. рис. 1. 16). Сначала рассмотрим участок, соответствующий промежутку $\langle x\rangle-\sigma\leqslant x\leqslant\langle x\rangle+\sigma$. Можно убедиться (пусть читатель поверит нам), что $\int\limits_{\langle x\rangle-\sigma}^{\langle x\rangle+\sigma} f(x)\,dx=0{,}683$.

Это означает, что вероятность попадания x в промежуток значений от $\langle x\rangle-\sigma$ до $\langle x\rangle+\sigma$ равна 0,683. Далее можно показать, что вероятность попадания в промежуток от $\langle x\rangle-2\sigma$ до $\langle x\rangle+2\sigma$ равна 0,954, а в промежуток от $\langle x\rangle-3\sigma$ до $\langle x\rangle+3\sigma$ равна 0,997. Таким образом, значения непрерывной случайной величины, подчиняющейся нормальному распределению, попадают в интервал $\langle x\rangle-3\sigma\leqslant x\leqslant\langle x\rangle+3\sigma$ с вероятностью 0,997. Такая вероятность практически равна единице. Поэтому на практике можно полагать, что фактически все значения рассматриваемой случайной величины находятся в пределах промежутка, простирающегося на 3σ вправо и на 3σ влево от $x=\langle x\rangle$. Это и есть «правило трех сигм».

ГЛАВА 2 Принятие решения

Потребности практики вызвали к жизни специальные научные методы, которые удобно объединять под названием «исследование операций». Под этим термином мы будем понимать применение математических, количественных методов для обоснования решений во всех областях целенаправленной человеческой деятельности.

Е. С. Вентцель

Трудности принятия решения

Принятие решения в условиях неопределенности. В жизни нам часто приходится принимать решения в условиях неопределенности. Неопределенность условий, в которых надо принимать решение, всегда в той или иной мере уменьшает нашу решительность. Вот простой пример. Куда поехать в очередной отпуск или во время каникул? Всем нам не раз приходилось размышлять над этим вопросом. Мы пытались предусмотреть различного рода неопределенности, касающиеся капризов погоды, условий проживания, возможных развлечений и т. д. В таких случаях мы стараемся отыскать наиболее удачный вариант, исходя из собственного опыта и советов знакомых, часто действуем «по вдохновению». В ситуации, которая касается только лично нас и наших близких, подобный *субъективный* подход к принятию решения вполне оправдан. Однако существует немало ситуаций, когда принятие того или иного решения затрагивает интересы большого количества людей и поэтому требует не субъективного, а *научного, математически аргументированного* подхода.

Известно, например, что общество не может функционировать без запасов продуктов, сырья, электроэнергии и т. п. Запасы создаются везде — на любом предприятии, в магазине, больнице, на транспорте. Но сколько необходимо тех или иных запасов в том или ином конкретном случае? Ясно, что их не должно быть слишком мало, иначе возникнут перебои в работе. Ясно также, что их не должно быть и слишком много, иначе они лягут на экономику тяжелым бременем, будут омертвленным капиталом. *Проблема запасов* — проблема исключительной важности. Она весьма сложна, поскольку решать ее практически всегда приходится в условиях неопределенности.

Два вида неопределенностей. Как же следует принимать решение в условиях неопределенности? Прежде всего надо выделить неизвестные факторы, обусловливающие неопределенность, оценить природу этих факторов. Различают два вида неопределенностей. Неопределенности первого вида обусловлены факторами, которые являются предметом изучения в теории вероятностей. Такие фак-

торы представляют собой либо *случайные величины*, либо *случайные функции*. Они описываются определенными статистическими характеристиками (например, математическим ожиданием и дисперсией), которые известны или же могут быть получены к нужному сроку. Такого рода неопределенности называют *вероятностными* или, иначе, *стохастическими*. Неопределенности второго вида обусловлены неизвестными факторами, которые нельзя отнести к категории случайных величин (случайных функций) по той причине, что набор реализаций этих факторов не обладает статистической устойчивостью и поэтому не позволяет ввести понятие вероятности. Такие неопределенности будем условно называть «плохими».

«Но позвольте,— может заметить читатель,— получается, что не всякое событие, которое нельзя точно предсказать, может быть отнесено к случайным событиям!» — «Да, не всякое»,— ответим мы. В предыдущей главе обсуждались случайные события, случайные величины, случайные функции. При этом неоднократно подчеркивалось, что рассматриваемая картина должна обладать *статистической устойчивостью*, которая как раз и выражается через вероятность. Однако возможны события, которые происходят от случая к случаю и в то же время никакой статистической устойчивости не обнаруживают. К таким событиям понятие вероятности неприменимо, соответственно неприменим и термин «случайный». Нельзя, например, говорить о вероятности получения двойки конкретным учеником по конкретному предмету. Ведь даже чисто умозрительно нельзя составить набора однотипных испытаний, имеющих в качестве одного из исходов данное событие. Не имеет смысла проводить такие испытания с набором учеников, так как у каждого ученика свои способности, своя степень подготовленности. Подобные испытания нельзя повторить и с одним учеником, так как от одного опроса к другому он будет, очевидно, все лучше и лучше ориентироваться в обстановке. Нельзя говорить о вероятности того или иного исхода встречи двух шахматистов одинакового ранга. Во всех подобных ситуациях нет набора однородных испытаний, которые позволили бы выявить выражаемую вероятностью устойчивость. Во всех таких ситуациях мы имеем дело с «плохой» неопределенностью.

К сожалению, в обыденной жизни мы не задумываемся над мудреным понятием «статистическая устойчивость» и щедро употребляем выражения «маловероятно», «вероятно», «вероятнее всего», «по всей вероятности» и тому подобные, применяя их, в частности, и к таким явлениям, которые никакой вероятностью не характеризуются. Мы склонны приписывать вероятностную природу всем событиям, которые не можем точно предсказать. Недаром же в начале нашего века возникла необходимость в уточнении понятия вероятности. Как уже отмечалось, это уточнение выразилось в разработке А. Н. Колмогоровым аксиоматического определения вероятности.

Элементы решения и показатель эффективности. Когда мы говорим о выборе решения, мы предполагаем, что возможны различ-

иые варнаиты поведеиия. Их называют *элементами решения*. Подчеркием, что в большннстве практически важных задач чнсло элемеитов решення весьма велико. Условимся обозначать через X множество элементов решеиия в рассматриваемой ситуации. Приятие решеиия озиачает, что мы выбираем нз этого множества какой-то элемеит x. Как определнть, какой из элемеитов решеиня будет иаиболее предпочтительным, иаиболее эффективным? Необходнм количествеииый крнтерий, позволяющий сравиивать разиые элемеиты решеиия по эффективиости. Будем называть такой критернй *показателем эффективности*. Этот показатель выбнрают, учнтывая ту *цель*, которая преследуется в даниой ситуацни: без опоздаиня явиться в школу, правнльно н быстро решить задачу, успеть сходнть в кино и т. д. Врач стремнтся иайти наиболее эффективиый метод лечеиия своего пацнента. Директор предпрнятня отвечает за выполиеиие плаиа выпуска продукции. Нанболее эффектнвным является тот элемент решення, который в нанбольшей степенн способствует достижеиню целн.

Предположим, что мы заиимаемся распродажей товаров. Наша цель — получить иаибольший доход. В качестве показателя эффектнвиости мы выбираем прнбыль и стремимся к тому, чтобы этот показатель оказался по возможиости иаибольшнм. В данном прнмере выбор показателя эффективности очевидеи. Существуют, одиако, более сложиые снтуацни, когда одиовременно преследуется несколько целей, иапрнмер увелнчеиие прнбылн, умеиьшеиие времеии распродажи, распределение товаров среди большего числа покупателей. В подобиых случаях приходнтся прнбегать к нескольким показателям эффективностн; такие задачи иазывают миогокритериальиыми.

Пусть W — показатель эффектнвиости (и прнтом едииствеииый). Казалось бы, проблема сводится к тому, что иадо найти такой элемеит решеиия x, при котором показатель W оказался бы максимальиым (нли, иапротнв, мниимальиым). Одиако следует помнить, что принятне решеиия происходит в условиях иеопределеиностн. Существуют иеизвестиые (случайные) факторы (обозначнм их через ξ), которые оказывают воздействие иа коиечиый результат и, следовательио, влияют иа показатель эффективиости W. Кроме того, всегда нмеется совокупность задаиных, зараиее нзвестиых факторов (обозиачим нх через α). Таким образом, показатель эффективиости оказывается зависящим от трех групп факторов — известиых факторов α, иеизвестиых (случайных) факторов ξ и выбраииого элемента решеиия x:

$$W = W(\alpha, \xi, x).$$

В примере с распродажей товаров под α иадо поинмать выделенные средства на прнобретеиие товаров, предоставлеиные помещення, сезои года и т. д. Под ξ поиимается количество покупателей в день (оио колеблется случайиым образом от одного дня к другому), время прихода покупателей (возможны случайиые скоплеиия покупателей, приводящие к длииным очередям), выбор

покупателями тех или иных товаров (спрос на данный товар случайно колеблется во времени) и т. д.
Поскольку факторы ξ случайные, то и показатель эффективности W оказывается случайной величиной. Возникает вопрос: а можно ли максимизировать (минимизировать) случайную величину? Ответ вполне ясен: разумеется, нельзя. Какой бы элемент решения x мы ни выбрали, величина W остается случайной величиной и нельзя заставить ее принять максимальное или минимальное значение. Такой ответ не должен обескураживать читателя. В условиях неопределенности мы действительно не можем со стопроцентной гарантией сделать показатель эффективности максимальным (минимальным). Однако соответствующим выбором элемента решения мы можем обеспечить это с *большой вероятностью*. Вот тут мы и подходим вплотную к приемам, используемым при принятии решения в условиях стохастической неопределенности.
Замена случайных факторов их средними значениями. Наиболее простой прием состоит в том, что случайные факторы ξ попросту заменяются их математическими ожиданиями. В результате задача становится строго детерминированной, показатель эффективности W может быть точно рассчитан и, в частности, может быть либо максимизирован, либо минимизирован. Такой прием широко используется при решении различных задач в физике и технике. Почти все используемые в этих задачах параметры (температура, разность потенциалов, освещенность, давление и т. д.) являются, строго говоря, случайными функциями. Как правило, мы пренебрегаем случайностью физических параметров и при решении различных задач пользуемся их средними значениями.
Такой прием оправдан, если отклонения параметров от их средних значений незначительны. Он не годится, если влияние случайностей на интересующий нас исход существенно. Например, при организации работы авторемонтной мастерской принципиально нельзя пренебрегать случайностью моментов возникновения неисправностей у автомашин, а также случайностью характера самих неисправностей и случайностью времени выполнения ремонтных операций. При рассмотрении шумов в электронной аппаратуре принципиально нельзя пренебрегать случайностями в поведении электронных потоков. В подобных примерах факторы ξ выступают как *существенно случайные* факторы.
Оптимизация в среднем. Если факторы ξ являются существенно случайными, то можно воспользоваться приемом, называемым *оптимизацией в среднем*. Он состоит в том, что в качестве показателя эффективности рассматривается не сама случайная величина W, а ее математическое ожидание $\langle W \rangle$, которое и пытаются максимизировать или минимизировать.
Разумеется, при таком подходе неопределенность сохраняется. Эффективность того или иного элемента решения x для конкретных значений случайных параметров ξ может оказаться существенно отличающейся от ожидаемой. Если мы оптимизируем операцию в среднем, то можем быть уверенными, что после многих

повторений операции в итоге обязательно будем иметь выигрыш. Следует иметь в виду, что оптимизация в среднем допустима лишь тогда, когда выигрыши повторяемых операций *суммируются*, так что «минусы» в одних операциях могут компенсироваться «плюсами» в других. Так, оптимизация в среднем оправдана, когда стремятся повысить прибыль, получаемую, например, при распродаже товаров. Прибыли, получаемые в разные дни, суммируются, так что случайно возникающие «неудачные» дни могут быть скомпенсированы «удачными» днями.

Но вот иной пример. Предположим, что рассматривается эффективность работы службы неотложной медицинской помощи большого города. В качестве показателя эффективности выберем время ожидания врача по вызову. Это время желательно сделать минимальным. В данном случае нельзя применить оптимизацию в среднем, поскольку слишком долгое ожидание врача одним больным отнюдь не компенсируется быстрым обслуживанием другого больного.

Стохастическое ограничение. Выдвинем дополнительное требование. Пусть время ожидания врача W будет меньше некоторого значения W_0. Так как W случайная величина, то нельзя просто потребовать выполнения неравенства $W < W_0$. Можно лишь потребовать, чтобы это неравенство выполнялось с достаточно большой вероятностью, например с вероятностью не меньше 0,99. Учет этого требования означает, что из множества X должны быть изъяты те элементы x, для которых рассматриваемое требование не удовлетворяется. Подобные ограничения называют *стохастическими ограничениями*. Естественно, что использование таких ограничений заметно усложняет проблему принятия решения.

Случайные процессы с дискретными состояниями

Под случайным процессом понимают процесс перехода системы из одних состояний в другие, протекающий случайным образом. В данной главе мы будем рассматривать случайные процессы с *дискретными состояниями*. Предполагаем, что система характеризуется набором дискретных состояний (конечным либо бесконечным). Случайные переходы системы между этими состояниями имеют характер *мгновенных скачков*.

Граф состояний. Случайные процессы с дискретными состояниями удобно рассматривать, используя схему, называемую *графом состояний*. На схеме условно изображают возможные состояния системы и показывают (при помощи стрелок) возможные переходы между состояниями.

Рассмотрим пример. Пусть система состоит из двух устройств, каждое из которых производит одну и ту же продукцию. Устройства в ходе работы могут выйти из строя (отказать). Отказавшее устройство немедленно начинают ремонтировать. Рассматриваемая система имеет четыре состояния: S_1 — оба устройства работают, S_2 — первое устройство ремонтируется (после отказа), второе

работает, S_3— второе ремонтируется, первое работает, S_4— оба ремонтируются. Граф состояний представлен на рисунке 2.1. Переходы $S_1 \to S_2$, $S_1 \to S_3$, $S_2 \to S_4$, $S_3 \to S_4$ совершаются в результате происходящих в системе отказов. Обратные переходы являются следствием ремонтных работ. Отказы устройств происходят в непредсказуемые моменты времени. Случайны также моменты времени, соответствующие окончанию ремонтов. Поэтому показанные стрелками на рисунке процессы перехода системы из одних состояний в другие являются случайными процессами.

Заметим, что на рисунке не показаны переходы $S_1 \to S_4$ и $S_4 \to S_1$. Первый переход отвечает одновременному отказу обоих устройств, а второй — одновременному окончанию ремонта обоих устройств. Можно полагать, что вероятность таких совпадений равна нулю.

Поток событий. Предположим, что однородные события следуют одно за другим в случайные моменты времени. Будем говорить о *потоке событий.* Это может быть поток заказов такси по телефону, поток включений приборов в бытовой электросети, поток сбоев в работе некоторого устройства и т. д.

Предположим, что диспетчер таксомоторного парка фиксирует моменты поступления заказов такси в течение некоторого промежутка времени, например от 12.00 до 14.00. Обозначив эти моменты в виде точек на оси времени, диспетчер получает картину, показанную на рисунке 2.2, *а*. Она изображает условно одну из реализаций потока заказов такси для рассматриваемого промежутка времени. Еще три реализации такого потока событий показаны на рисунках 2.2, *б, в, г*; они зафиксированы в другие дни. Моменты появления событий в каждой реализации потока случайны. В то же время поток событий обнаруживает статистическую устойчивость: полное количество событий на рассматриваемом промежутке времени слабо меняется от эксперимента к эксперименту (от одной реализации потока к другой). Можно видеть, что числа событий в представленных реализациях потока равны 19, 20, 21, 18.

В предыдущей главе под случайным событием понимался тот или иной исход опыта, характеризующийся определенной вероятностью. При рассмотрении потока событий термин «событие» имеет иное значение. Нет смысла говорить о вероятности того или иного исхода (события), поскольку все события рассматриваются как однотипные, не отличающиеся одно от другого. Так, любой заказ такси есть одна из заявок, сама по себе она ничем не отличается от остальных заявок. Теперь рассматриваются иные вероятности, например вероятности того, что на некотором определенном промежутке времени (предположим, на промежутке от t до $t+\Delta t$, выделенном на рисунке) событие появится ровно один раз, два раза, три раза и т. д.

Понятие «поток событий» применяется при рассмотрении случайных процессов в системах с дискретными состояниями. При этом полагают, что переходы системы из одних состояний в другие происходят под действием соответствующих потоков событий. Как только в потоке появляется событие, тут же совершается мгно-

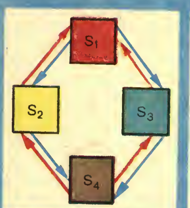


Рис. 2.1

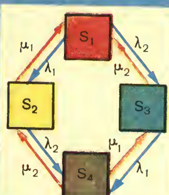


Рис. 2.3

а
б
в
г
t
t+Δt

Рис. 2.2

венный переход. В примере с графом состояний, изображенным на рисунке 2.1, переходы $S_1 \to S_2$ и $S_3 \to S_4$ совершаются под действием потока событий, представляющего собой поток отказов первого устройства. Переходы $S_1 \to S_3$ и $S_2 \to S_4$ совершаются под действием потока отказов второго устройства. Обратные переходы вызываются потоком событий, в качестве которых выступают «окончания ремонтов»: переходы $S_2 \to S_1$ и $S_4 \to S_3$ — потоком окончаний ремонтов первого устройства, а переходы $S_3 \to S_1$ и $S_4 \to S_2$ — потоком окончаний ремонтов второго устройства.

Переход системы из состояния S_i в состояние S_j происходит, как только в соответствующем потоке появляется очередное событие. Напрашивается вывод, что вероятность перехода $S_i \to S_j$ в некоторый выбранный момент времени t должна равняться вероятности появления события в потоке в данный момент времени. Не имеет смысла говорить о вероятности перехода в конкретный момент t. Как и вероятность любого конкретного значения непрерывной случайной величины, эта вероятность равна нулю. Здесь сказывается непрерывность времени. Поэтому говорят о вероятности перехода (вероятности появления события в потоке) не в момент t, а на промежутке времени от t до $t + \Delta t$. Будем обозначать эту вероятность через $P_{ij}(t;\ \Delta t)$. В пределе при $\Delta t \to 0$ мы приходим к понятию *плотности вероятности перехода* в момент t:

$$\lambda_{ij}(t) = \lim_{\Delta t \to 0} \frac{P_{ij}(t;\ \Delta t)}{\Delta t}\ . \tag{2.1}$$

Эту величину называют также *интенсивностью потока событий*, вызывающих рассматриваемый переход.

В общем случае интенсивность потока зависит от времени. Важно помнить, что зависимость интенсивности потока от времени не связана с расположением сгущений и разрежений событий в той или иной реализации потока. В дальнейшем для простоты будем полагать, что плотность вероятности переходов и, следовательно, интенсивность потоков событий от времени не зависит, т. е. будем рассматривать *стационарные* потоки.

Уравнения Колмогорова для стационарного режима. Обозначим через p_i вероятность того, что система находится в состоянии S_i (ограничимся рассмотрением *стационарного* режима — когда вероятности p_i не зависят от времени). В качестве примера выберем систему, граф состояний которой дан на рисунке 2.1. Пусть λ_1 — интенсивность потока отказов первого устройства, а λ_2 — второго устройства; μ_1 — интенсивность потока окончаний ремонтов первого устройства, а μ_2 — второго устройства. Учитывая интенсивности потоков событий, получаем *размеченный* граф состояний; он изображен на рисунке 2.3.

Представим себе, что имеется N одинаковых систем, описываемых графом состояний, изображенным на рисунке 2.3. Пусть $N \gg 1$. Число систем, находящихся в состоянии S_i, равно Np_i (это утверждение тем точнее, чем больше N). Рассмотрим конкретное состояние, например S_1. Из этого состояния возможны переходы в со-

стояния S_2 и S_3 — с суммарной вероятностью $\lambda_1+\lambda_2$, отнесенной к единице времени. (В стационарном случае плотность вероятности есть вероятность за конечный промежуток времени Δt, деленная на Δt.) Таким образом, число уходов из состояния S_1 в единицу времени в рассматриваемом коллективе систем равно $Np_1(\lambda_1+\lambda_2)$. Здесь просматривается общее правило: совершаемое в единицу времени число переходов $S_i \to S_j$ равно произведению числа систем в состоянии S_i (в исходном состоянии) на вероятность перехода, отнесенную к единице времени. Мы рассмотрели уходы из состояния S_1. Приходы в это состояние совершаются из S_2 и S_3. Число приходов в S_1 в единицу времени равно $Np_2\mu_1+Np_3\mu_2$. Поскольку рассматривается стационарный режим, то числа уходов и приходов для каждого состояния должны быть сбалансированы. Следовательно,

$$Np_1(\lambda_1+\lambda_2)=Np_2\mu_1+Np_3\mu_2.$$

Рассматривая баланс уходов и приходов для каждого из четырех состояний и сокращая в уравнениях общий множитель N, получаем следующие уравнения относительно вероятностей p_1, p_2, p_3, p_4:

для состояния S_1: $(\lambda_1+\lambda_2)p_1=\mu_1p_2+\mu_2p_3$;
для состояния S_2: $(\lambda_2+\mu_1)p_2=\lambda_1p_1+\mu_2p_4$;
для состояния S_3: $(\lambda_1+\mu_2)p_3=\lambda_2p_1+\mu_1p_4$;
для состояния S_4: $(\mu_1+\mu_2)p_4=\lambda_2p_2+\lambda_1p_3$.

Нетрудно убедиться, что четвертое уравнение может быть получено сложением первых трех. Вместо этого уравнения воспользуемся уравнением

$$p_1+p_2+p_3+p_4=1,$$

которое означает, что система с достоверностью находится в каком-либо из четырех состояний. Таким образом, приходим к системе уравнений:

$$\left.\begin{array}{l}(\lambda_1+\lambda_2)p_1=\mu_1p_2+\mu_2p_3;\\(\lambda_2+\mu_1)p_2=\lambda_1p_1+\mu_2p_4;\\(\lambda_1+\mu_2)p_3=\lambda_2p_1+\mu_1p_4;\\p_1+p_2+p_3+p_4=1.\end{array}\right\} \qquad (2.2)$$

Это есть *уравнения Колмогорова*, записанные для системы, граф состояний которой показан на рисунке 2.3.

Какую рационализацию следует выбрать? Проанализируем конкретную ситуацию, используя уравнения (2.2). Соответствующий этим уравнениям граф состояний (см. рис. 2.3) описывает систему, которая, как мы условились, состоит из двух устройств, производящих некоторую продукцию. Предположим, что второе устройство в данной системе является более современным и имеет производительность вдвое более высокую, чем первое устройство. Первое устройство приносит в единицу времени доход, равный 5 условным единицам, а второе — 10 единицам. К сожалению, отказы второго устройства происходят в среднем вдвое чаще, чем первого; по-

ложим $\lambda_1=1$, $\lambda_2=2$. Интенсивности потоков окончаний ремонтов примем равными $\mu_1=2$, $\mu_2=3$. Используя заданные интенсивности потоков отказов и потоков окончаний ремонтов, перепишем (2.2) в виде:

$$\left.\begin{aligned} &3p_1=2p_2+3p_3;\\ &4p_2=p_1+3p_4;\\ &4p_3=2p_1+2p_4;\\ &p_1+p_2+p_3+p_4=1.\end{aligned}\right\}$$

Решая эту систему уравнений, находим: $p_1=0{,}4$; $p_2=0{,}2$; $p_3=0{,}27$; $p_4=0{,}13$. Это означает, что в среднем 40% времени оба устройства работают одновременно (состояние S_1 на рисунке), 20% времени работает только первое устройство, а второе при этом ремонтируется (состояние S_2), 27% времени работает только второе устройство, а первое при этом ремонтируется (состояние S_3), 13% времени оба устройства одновременно находятся в состоянии ремонта (состояние S_4). Нетрудно подсчитать доход, который дает система из двух рассматриваемых устройств в единицу времени: $(5+10)\times0{,}4+5\times0{,}2+10\times0{,}27=9{,}7$ усл. единиц.

Предположим, что предлагается некоторая рационализация, позволяющая вдвое сократить время ремонта либо первого, либо второго устройства. По ряду причин мы можем применить рационализацию лишь к одному из устройств. Спрашивается, какое устройство следует выбрать, первое или второе? Вот конкретный пример практической ситуации, когда, пользуясь теорией вероятностей, надо научно обосновать принятие решения.

Допустим, что мы выбираем первое устройство. В результате рационализации интенсивность потока окончаний ремонтов этого устройства увеличивается вдвое, так что теперь $\mu_1=4$ (остальные интенсивности остаются прежними: $\lambda_1=1$, $\lambda_2=2$, $\mu_2=3$). Уравнения (2.2) принимают теперь следующий вид:

$$\left.\begin{aligned} &3p_1=4p_2+3p_3;\\ &6p_2=p_1+3p_4;\\ &4p_3=2p_1+4p_4;\\ &p_1+p_2+p_3+p_4=1.\end{aligned}\right\}$$

Решая эту систему, находим: $p_1=0{,}48$; $p_2=0{,}12$; $p_3=0{,}32$; $p_4=0{,}08$. С учетом полученных вероятностей определяем доход, который теперь будет давать рассматриваемая система: $(5+10)\times0{,}48+$ $+5\times0{,}12+10\times0{,}32=11$ усл. единиц.

Если мы выберем второе устройство, то в результате рационализации удвоится интенсивность μ_2. В этом случае: $\lambda_1=1$, $\lambda_2=2$, $\mu_1=2$, $\mu_2=6$. Уравнения (2.2) принимают вид:

$$\left.\begin{aligned} &3p_1=2p_2+6p_3;\\ &4p_2=p_1+6p_4;\\ &7p_3=2p_1+2p_4;\\ &p_1+p_2+p_3+p_4=1.\end{aligned}\right\}$$

Решая эту систему, находим: $p_1=0{,}5$; $p_2=0{,}25$; $p_3=0{,}17$; $p_4=0{,}08$.

Подсчитываем доход: $(5+10)\times0,5+5\times0,25+10\times0,17=10,45$ усл. единиц. Мы видим, таким образом, что выгоднее применять рационализацию к первому устройству.

Системы массового обслуживания

Проблема массового обслуживания. Современное общество не может обойтись без разветвленной сети *систем массового обслуживания.* К таким системам относятся телефонные станции, магазины, поликлиники, предприятия общественного питания, билетные кассы, автозаправочные станции, парикмахерские и т. п. Несмотря на разнообразие, все эти системы имеют общие черты и общие проблемы.

Одной из таких общих проблем является проблема очередей. Когда мы приходим в поликлинику, столовую, магазин, парикмахерскую, билетную кассу, нам, как правило, всегда приходится встать в очередь. Очереди подстерегают нас буквально на каждом шагу, отнимая время и нарушая планы.

Ясно, что источник подобных проблем лежит в *случайном характере* явлений, происходящих в системах массового обслуживания. Случаен поток вызовов, поступающих на телефонную станцию, случайна длительность телефонного разговора. Случайности принципиально нельзя устранить. Но их можно должным образом учесть и, как следствие, можно достаточно рационально организовать систему массового обслуживания. Подобные вопросы начали исследоваться в первой четверти нашего столетия. Были сформулированы и рассмотрены математические задачи, в которых моделировались случайные процессы в системах с дискретными состояниями. Возникло и стало развиваться новое направление в теории вероятностей, названное по предложению видного советского математика А. Я. Хинчина «теорией массового обслуживания».

Исторически эта теория берет начало от работ, в которых рассматривалась проблема перегрузки телефонных линий, весьма обострившаяся в начале столетия. Начальный период в развитии теории связан с работами известного датского ученого А. Эрланга, относящимися к 1908—1922 годам. С этого времени интерес к проблемам массового обслуживания быстро нарастает. Стремление более рационально обслуживать большие массы людей приводило к необходимости изучения закономерностей образования очередей. Очень скоро стало ясно, что задачи, рассматриваемые в теории массового обслуживания, выходят за рамки сферы обслуживания, имеют более широкую область применения. Предположим, что рабочий работает на нескольких станках. В случайные моменты времени в станках возникают неисправности, требующие срочного вмешательства. Длительность операции по устранению неисправности — случайная величина. В результате возникает такая же ситуация, как и в обычных системах массового обслуживания. Только здесь речь идет уже не об обслуживании многих людей

некоторой системой, а об обслуживании многих станков одним рабочим.

Диапазон практических задач теории массового обслуживания необычайно широк. Мы обращаемся к этой теории, когда желаем организовать эффективную работу современного морского порта, когда, в частности, анализируем пропускную способность крупного причала. Мы обращаемся к теории массового обслуживания и тогда, когда рассматриваем работу счетчика Гейгера — Мюллера. Такие счетчики применяются в ядерной физике, они предназначены для счета частиц. Когда частица попадает внутрь счетчика, там возникает разряд, по которому и фиксируется появление частицы. Пока происходит разряд, новая частица уже не может быть зарегистрирована («обслужена») данным счетчиком. Случайны моменты появления частицы в счетчике, случайные колебания испытывает длительность разряда (время «обслуживания»). Налицо ситуация, характерная для систем массового обслуживания.

Основные понятия. Любая система массового обслуживания предназначена для выполнения некоторого *потока заявок*. В роли заявки может выступать появление пассажира в билетной кассе, возникновение неисправности в устройстве, приход судна в порт, появление частицы в счетчике Гейгера — Мюллера. Система имеет одну или несколько обслуживающих единиц, их принято называть *каналами обслуживания*. Когда вы, приходя в парикмахерскую, интересуетесь количеством работающих мастеров, вы тем самым выясняете, чему равно в данном случае число каналов обслуживания. В других ситуациях это есть число кассиров в билетной кассе, число телефонов на переговорном пункте, число причалов в порту, число колонок на автозаправочной станции и т. д. Приходя в поликлинику к определенному врачу, мы имеем дело с одноканальной системой массового обслуживания.

Рассматривая работу той или иной системы массового обслуживания, надо учитывать прежде всего число каналов обслуживания, число заявок, поступающих на вход системы в единицу времени, длительность обслуживания заявки. Существенно, что число поступающих в систему заявок, моменты их поступления, длительность обслуживания заявки относятся, как правило, к категории *случайных* факторов. Поэтому теория массового обслуживания должна рассматриваться в рамках *теории случайных процессов*.

Случайные процессы такого типа обсуждались в предыдущем параграфе. Имеются в виду случайные процессы с *дискретными состояниями*. Переходы системы из одних состояний в другие происходят под действием потока заявок, поступающих на вход системы, и *потока обслуживаний*. Под последним понимается поток заявок, обслуживаемых одним непрерывно занятым каналом системы.

Виды систем массового обслуживания. Системы массового обслуживания бывают двух видов: *системы с отказами* и *системы с очередью*. Если заявка поступает в систему с отказами в момент времени, когда все каналы заняты, то она получает «отказ» и выходит из игры. С такими системами мы знакомы на примере

телефона. Если абонент занят, ваша заявка получает отказ, и вы можете повесить трубку. Повторно набирая телефонный номер, вы тем самым посылаете новую заявку.

Наиболее часто на практике встречаются системы с очередью, или, иначе, системы с ожиданием. Недаром теорию массового обслуживания называют также *теорией очередей*. В таких системах заявка, появившаяся в момент, когда все каналы обслуживания заняты, встает в *очередь* и ожидает, пока не освободится один из каналов. Существуют системы с *неограниченной очередью* (стоящая в очереди заявка рано или поздно будет обслужена, при этом число мест в очереди не ограничено) и системы с *ограниченной очередью*. Ограничения могут быть разными — по числу заявок, одновременно стоящих в очереди (в очереди должно быть не больше некоторого числа заявок, всякая дополнительная заявка получает отказ); по времени пребывания заявки в очереди (после некоторого срока пребывания в очереди заявка, если она не начала обслуживаться, покидает очередь); по времени работы системы (прием заявок к обслуживанию происходит в течение определенного времени) и т. д.

Учитывается также дисциплина обслуживания. Обычно заявки обслуживаются в порядке их поступления в систему. Но возможно также *обслуживание с приоритетом*, когда некоторые заявки обслуживаются вне очереди. При этом заявка с более высоким приоритетом, поступив в систему, может оборвать уже начавшееся обслуживание заявки с меньшим приоритетом, а может дождаться окончания предыдущего обслуживания. В первом случае говорят об *абсолютном* приоритете, а во втором — об *относительном*.

Системы массового обслуживания всегда являются *многокритериальными*; они характеризуются *набором* показателей эффективности. В качестве таковых могут выступать среднее число заявок, которое обслуживает система в единицу времени; среднее число занятых каналов обслуживания; среднее число заявок, находящихся в очереди; среднее время ожидания обслуживания; средний процент заявок, получающих отказ; вероятность того, что поступившая в систему заявка будет немедленно принята к обслуживанию. Возможны также другие показатели эффективности.

Представляется вполне естественным, что при организации работы той или иной системы массового обслуживания следует стремиться к сокращению среднего числа заявок, находящихся в очереди, к сокращению времени ожидания обслуживания. Хотелось бы сделать максимальной вероятность того, что поступившая в систему заявка будет немедленно принята к обслуживанию, а средний процент заявок, получающих отказ, хотелось бы свести к минимуму. Для этого надо увеличивать производительность системы (уменьшать время обслуживания заявки), рационализировать режим работы системы, увеличивать число каналов обслуживания. Однако при увеличении числа каналов неизбежно уменьшается такой показатель, как среднее число занятых каналов. Это означает, что возрастает время, в течение которого тот или иной канал не будет

заият обслуживаиием, будет простаивать. В результате сиизится эффективиость использоваиия системы. Таким образом, возиикает иеобходимость *оптимизации* работы системы. Число каиалов обслуживаиия ие должио быть чрезмерио малым (чтобы ие возиикали длииные очереди и ие возрастало число отказов), ио оио ие должио быть и слишком большим (чтобы ие росли число и длительиость простоев в каиалах обслуживания).

Системы с отказами. Простейший тип системы массового обслуживания — *одноканальная система с отказами.* В качестве примера такой системы можио указать систему, состоящую из одиой телефоииой лииии, или детектор частиц, состоящий из одиого счетчика Гейгера — Мюллера. Граф состояиий рассматриваемой системы показаи иа рисуике 2.4, *а.* Здесь состояиие S_0— каиал свободеи, состояиие S_1— каиал заият. Через λ обозиачеиа иитеисивиость потока заявок, а через μ — иитеисивиость потока обслуживаиий. Этот граф состояиий очеиь прост. Если система иаходится в состояиии S_0, то поступающая иа ее вход заявка переводит систему в состояиие S_1; иачинается обслуживаиие. Как только обслуживаиие закаичивается, система возвращается в состояиие S_0 и готова приият ь иовую заявку.

Не остаиавливаясь подробиее иа даииом типе систем, перейдем к более общему случаю — *n-канальной системе с отказами.* Примером может служить система, состоящая из n телефоииых лииий. Имеиио такую систему рассматривал в свое время осиователь теории массового обслуживаиия Эрлаиг. Соответствующий граф состояиий даи иа рисуике 2.4, *б.* Для состояиий системы использоваиы обозиачеиия: S_0— все каиалы свободиы, S_1— заият одии каиал, остальиые свободиы, S_2— заияты два каиала, остальиые свободиы... S_n— заияты все n каиалов. Как и в предыдущем примере, λ — иитеисивиость потока заявок, μ — иитеисивиость потока обслуживаиий.

Пусть система иаходится в состояиии S_0. Как только приходит заявка, одии из каиалов стаиовится занятым — система переходит в состояиие S_1. Если система иаходится в S_1 и приходит очередиая заявка, то стаиовятся заиятыми уже два каиала — система переходит из S_1 в S_2. Итак, одии и тот же поток (поток заявок с иитеисивиостью λ) переводит систему из любого состояиия в соседиее в иаправлеиии *слева направо* (см. граф состояиий, показаииый иа рисуике). Вопрос о потоках событий, приводящих к переходам между соседними состояииями в иаправлеиии *справа налево,* иемиого более сложен. Если система иаходится в S_1 (заият одии каиал), то очередное событие в потоке обслуживаиий освобождает этот каиал и переводит систему в состояиие S_0. Напомиим, что иитенсивиость потока обслуживаиий есть μ. Предположим теперь, что система иаходится в S_2, т. е. заияты два каиала. Средиее время обслуживаиия в каждом каиале одио и то же. Каждый каиал освобождается под действием одиого и того же потока обслуживаиий иитеисивиостью μ. Для перехода системы из S_2 в S_1 безразличио, какой имеиио из двух каиалов освободится.

Следовательно, поток событий, переводящий систему из S_2 в S_1, имеет интенсивность 2μ. Для перехода системы из S_3 в S_2 безразлично, какой именно из трех занятых каналов освободится. Поток событий, переводящий систему из S_3 в S_2, имеет интенсивность 3μ. И так далее. Легко сообразить, что поток событий, переводящий систему из S_k в S_{k-1}, имеет интенсивность $k\mu$.

Будем полагать, что система находится в стационарном режиме. Применяя правило, сформулированное в предыдущем параграфе, и пользуясь графом состояний, данным на рисунке 2.4, *б*, можно составить уравнения Колмогорова для вероятностей p_0, p_1, p_2, ..., p_n (напомним: p_i— вероятность того, что система находится в состоянии S_i). Получаем следующую систему уравнений:

$$\left.\begin{aligned}
&\lambda p_0 = \mu p_1;\\
&(\lambda+\mu)p_1 = \lambda p_0 + 2\mu p_2;\\
&(\lambda+2\mu)p_2 = \lambda p_1 + 3\mu p_3;\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
&(\lambda+k\mu)p_k = \lambda p_{k-1} + (k+1)\mu p_{k+1};\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
&[\lambda+(n-1)\mu]\, p_{n-1} = \lambda p_{n-2} + n\mu p_n;\\
&p_0 + p_1 + p_2 + \ldots + p_n = 1.
\end{aligned}\right\} \quad (2.3)$$

Эта система уравнений легко решается. Используя первое уравнение, выражаем p_1 через p_0 и подставляем во второе. Затем из второго уравнения выражаем p_2 через p_0 и подставляем в третье. И так далее. На предпоследнем этапе выражаем p_n через p_0. Наконец, полученные на каждом шаге результаты подставляем в последнее уравнение и находим выражение для p_0.

Итак,

$$p_0 = \left[1 + \lambda/\mu + \frac{(\lambda/\mu)^2}{2!} + \frac{(\lambda/\mu)^3}{3!} + \ldots + \frac{(\lambda/\mu)^n}{n!}\right]^{-1};$$

$$p_k = \frac{(\lambda/\mu)^k}{k!} p_0 \quad (k = 1,\ 2,\ 3,\ \ldots,\ n). \quad (2.4)$$

Заявка получает отказ, если приходит тогда, когда все n каналов обслуживания заняты, т. е. когда система находится в состоянии S_n. Вероятность того, что система находится в S_n, равна p_n. Это и есть вероятность того, что поступающая на вход системы заявка получает отказ и не обслуживается. Отсюда находим вероятность того, что поступившая в систему заявка будет принята к обслуживанию:

$$Q = 1 - p_n = 1 - \frac{(\lambda/\mu)^n}{n!} p_0. \quad (2.5)$$

Умножив Q на λ, получаем интенсивность потока обслуженных системой заявок. Так как каждый занятый канал обслуживает в единицу времени в среднем μ заявок, то, разделив полученное выше произведение на μ, находим среднее число занятых каналов в системе:

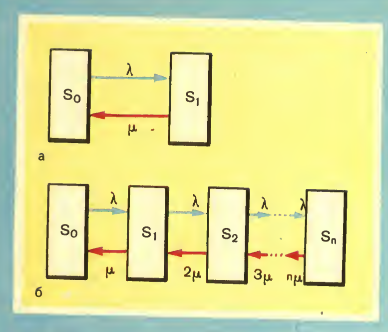
$$\langle N \rangle = \frac{\lambda}{\mu}\left[1 - \frac{(\lambda/\mu)^n}{n!} p_0\right]. \quad (2.6)$$

Сколько требуется каналов обслуживания? Рассмотрим конкретный пример. Предположим, что на станцию телефонного обслуживания поступают в среднем 1,5 заявки в минуту, а поток обслуживаний имеет интенсивность, равную 0,5 заявки в минуту (среднее время обслуживания одной заявки составляет две минуты). Таким образом, $\lambda/\mu=3$. Пусть станция имеет три канала обслуживания (три телефонных линии). Используя формулы (2.4) — (2.6) для $\lambda/\mu=3$ и $n=3$, можно рассчитать, что вероятность обслуживания поступившей заявки составляет всего лишь 65%. При этом среднее число занятых каналов равно 1,96, что составляет 65% от всего числа каналов. Итак, 35% поступающих в систему заявок получают отказ и к обслуживанию не принимаются. Это слишком много. Мы принимаем решение увеличить число каналов обслуживания. Попробуем добавить еще один, четвертый, канал. В этом случае вероятность обслуживания заявки возрастает до 79% (вероятность отказа уменьшается до 21%). Вместе с тем среднее число занятых каналов становится равным 2,38, что составляет 60% от всего числа каналов. По-видимому, решение о добавлении четвертого канала является вполне обоснованным, поскольку при сравнительно небольшом снижении процента занятых каналов (с 65 до 60%) происходит довольно существенное увеличение вероятности обслуживания — с 65 до 79%. Дальнейшее увеличение числа каналов может оказаться невыгодным из-за уменьшения эффективности использования системы вследствие простоев в каналах. Здесь необходим более детальный анализ с учетом стоимости каждого канала. Заметим, что при $n=5$ получаем $Q=89\%$, $\langle N\rangle/n=53\%$, а при $n=6$: $Q=94\%$, $\langle N\rangle/n=47\%$.

Одноканальная система с ограниченной очередью. Пусть ограничение осуществляется по числу заявок, стоящих в очереди; число мест в очереди равно m. Если все места заняты, то очередная заявка, поступающая в систему, получает отказ. Примером подобной системы может служить автозаправочная станция, имеющая одну колонку (один канал обслуживания) и площадку, на которой могут находиться одновременно не более m автомашин. Если все места на площадке заняты, то очередная машина, прибывшая к станции, не останавливается, а проезжает мимо.

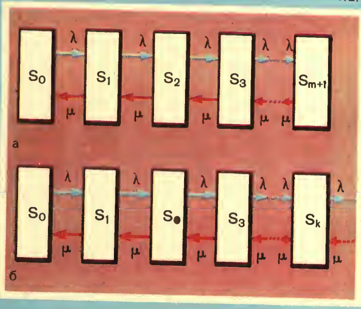
Граф состояний рассматриваемой системы показан на рисунке 2.5, *а*. Здесь: S_0— канал свободен, S_1— канал занят, S_2— канал занят, одна заявка стоит в очереди, S_3— канал занят, в очереди стоят две заявки... S_{m+1}— канал занят, в очереди стоят m заявок. Как обычно, λ — интенсивность потока заявок, μ — интенсивность потока обслуживаний. Уравнения Колмогорова для стационарного режима имеют в данном случае следующий вид:

$$\left.\begin{aligned} &\lambda p_0=\mu p_1;\\ &(\lambda+\mu)p_1=\lambda p_0+\mu p_2;\\ &\ldots\ldots\ldots\ldots\\ &(\lambda+\mu)p_m=\lambda p_{m-1}+\mu p_{m+1};\\ &p_0+p_1+p_2+\ldots+p_m+p_{m+1}=1. \end{aligned}\right\} \qquad (2.7)$$



S_0
S_1
λ
μ
а
S_0
S_1
S_2
S_n
λ
λ
λ
λ
μ
2μ
3μ
nμ
б

Рис. 2.4



S_0
S_1
S_2
S_3
S_{m+1}
λ
λ
λ
λ
λ
μ
μ
μ
μ
μ
а
S_0
S_1
$S_●$
S_3
S_k
λ
λ
λ
λ
λ
λ
μ
μ
μ
μ
μ
μ
б

Рис. 2.5

Решая эту систему и вводя обозначение $\varrho=\lambda/\mu$, получаем:

$$p_0=\frac{1}{1+\varrho+\varrho^2+\varrho^3+\ldots+\varrho^{m+1}}=\frac{1-\varrho}{1-\varrho^{m+2}}; \qquad p_k=\varrho^k p_0. \tag{2.8}$$

Заявка получает отказ, если приходит тогда, когда занят канал обслуживания и в очереди стоят m заявок, т. е. когда система находится в состоянии S_{m+1}. Таким образом, вероятность отказа есть p_{m+1}. Среднее число заявок в очереди определяется очевидным соотношением: $\langle r\rangle=\sum_{k=1}^{m} kp_{k+1}$ (p_{k+1} есть вероятность того, что в очереди стоят k заявок). Среднее время ожидания в очереди равно отношению $\langle r\rangle/\lambda$.

Предположим, что на автозаправочную станцию прибывает в минуту в среднем одна машина ($\lambda=1$ заявка в минуту) и что длительность заправки составляет в среднем 2 мин ($\mu=1/2$). Таким образом, $\varrho=\lambda/\mu=2$. Если число мест в очереди $m=3$, то, как нетрудно подсчитать, вероятность отказа составляет 51,6%, а среднее время ожидания в очереди равно 2,1 мин. Допустим, что, желая снизить вероятность отказа, мы попробуем увеличить вдвое число мест в очереди. Оказывается, что при $m=6$ вероятность отказа равна 50,2%, т. е. остается фактически прежней, зато заметно возрастает время ожидания в очереди, достигая теперь 5 мин. Из (2.8) видно, что если $\varrho>1$, то при больших m вероятность отказа стабилизируется, становясь равной $(\varrho-1)/\varrho$. Чтобы существенно снизить вероятность отказа, необходимо (если нельзя уменьшить ϱ) переходить к многоканальным системам.

Одноканальная система с неограниченной очередью. Заметим, что такие системы массового обслуживания довольно часто встречаются на практике: врач, принимающий больных, телефон-автомат с одной будкой, порт с одним причалом, где выгружаются прибывающие суда, и т. д. Граф состояний рассматриваемой системы представлен на рисунке 2.5, *б*. Здесь: S_0— канал свободен, S_1— канал занят, S_2— канал занят, в очереди стоит одна заявка, S_3— канал занят, в очереди стоят две заявки... S_k— канал занят, в очереди стоят $k-1$ заявок...

До сих пор мы рассматривали графы с конечным числом состояний. Здесь же мы встречаемся с системой, характеризующейся бесконечным числом дискретных состояний. Возникает вопрос: можно ли говорить о стационарном режиме для такой системы? Оказывается, можно. При этом необходимо, чтобы выполнялось неравенство $\varrho<1$. Если указанное неравенство выполняется, то сумму $1+\varrho+\varrho^2+$ $+\ldots+\varrho^{m+1}$ в (2.8) можно заменить суммой бесконечно убывающей геометрической прогрессии $1+\varrho+\varrho^2+\varrho^3+\ldots=1/(1-\varrho)$. В результате получаем

$$p_0=1-\varrho; \qquad p_k=\varrho^k p_0. \tag{2.9}$$

Если условие $\varrho<1$ не выполняется, то стационарный режим в рассматриваемой системе не устанавливается: очередь при $t\to\infty$ растет неограниченно.

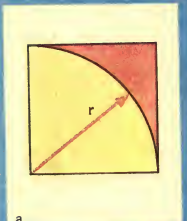
Метод статистических испытаний

Статистические испытания предполагают многократное повторение однотипных испытаний. Результат любого отдельного испытания случаен и сам по себе какого-либо интереса не представляет. В то же время совокупность большого числа подобных результатов оказывается весьма полезной. Она обнаруживает определенную устойчивость (ее называют *статистической устойчивостью*), которая позволяет количественно описать явление, исследуемое в данных испытаниях. Рассмотрим специальный метод исследования случайных процессов, основанный на статистических испытаниях. Его так и называют — *метод статистических испытаний*. Другое название этого метода, которое, кстати говоря, используется чаще,— *метод Монте-Карло*.

Заметим сразу же, что сам город Монте-Карло (столица княжества Монако), его жители и гости не имеют никакого отношения к рассматриваемому методу. Дело в том, что этот город широко известен своими игорными домами, в которых богатые туристы ставят на рулетку немалые суммы. Рулетка могла бы быть гербом этого города. Но ведь рулетка — это генератор случайных чисел. Именно это и имеют в виду, когда говорят о методе Монте-Карло.

Два примера, свидетельствующих о пользе статистических испытаний. *Первый пример*. На рисунке 2.6, *а* показан квадрат со стороной r, в который вписана четверть круга радиуса r. Отношение площади, закрашенной в желтый цвет, к площади квадрата равно $(\pi r^2)/4r^2=\pi/4$. Это отношение, а следовательно и число π, можно приближенно получить, если проделать следующие статистические испытания. Лист бумаги с рассматриваемым рисунком положим на горизонтальную поверхность и будем бросать на этот лист мелкие крупинки. Бросать надо без какого-либо прицеливания так, чтобы крупинка с равной вероятностью могла попасть в любую часть листа. Можно, например, завязать бросающему глаза. Разбросанные крупинки распределятся по поверхности листа случайным образом (рис. 2.6, *б*). Часть из них окажется вне квадрата — эти крупинки мы в дальнейшем не будем учитывать. Подсчитаем количество крупинок, попавших в квадрат (обозначим это число через N_1), и выделим те крупинки, которые оказались в пределах области, закрашенной в желтый цвет (N_2 крупинок). Поскольку крупинка имела одинаковую вероятность попасть в любой участок рисунка, то отношение N_2/N_1 при достаточно большом числе бросаний будет приближенно равняться отношению площади, закрашенной в желтый цвет, к площади квадрата, т. е. числу $\pi/4$. Это равенство будет тем точнее, чем больше число бросаний.

Данный пример интересен тем, что здесь определенное число (число π) отыскивается по результатам статистических испытаний. Можно сказать, что случайность использовалась здесь в качестве инструмента, при помощи которого был получен детерминистский результат — приближенное значение числа π.



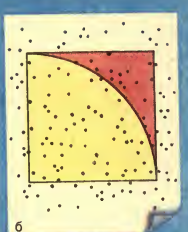


Рис. 2.6

y
1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0 0,1 0,2 0,3 0,4 0,5 0,6 0,7 0,8 0,9 1
x

Рис. 2.7

Второй пример. Значительно чаще статистические испытания используют для исследования *случайных событий, случайных процессов.* Предположим, что производится сборка изделия, состоящего из трех деталей (детали A, B, C). Перед сборщиком находятся три ящика — с деталями A, B и C соответственно. Пусть половина деталей каждого типа имеет размеры с положительными отклонениями от номинала, а половина — с отрицательными отклонениями. Изделие не может нормально функционировать лишь в тех случаях, когда все три детали имеют положительные отклонения. Сборщик берет детали из ящиков наугад. Спрашивается, какова вероятность сборки нормально функционирующего изделия?

Конечно, этот пример довольно прост. Искомую вероятность легко рассчитать. Вероятность получения бракованного изделия есть вероятность того, что все три детали окажутся с положительными отклонениями; она равна $\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=\frac{1}{8}$. Следовательно, вероятность сборки небракованного изделия есть $1-\frac{1}{8}=0{,}875$.

Забудем на время, что мы умеем вычислять вероятности. Воспользуемся статистическими испытаниями. В качестве таковых надо выбрать испытания, каждое из которых имеет два равновероятных исхода, например подбрасывание монеты. Возьмем три монеты — A, B, C. Каждая монета сопоставляется с соответствующей деталью, используемой при сборке изделия. Выпадение герба при подбрасывании монеты будет означать, что соответствующая деталь имеет в данном испытании положительное отклонение, а выпадение «решки» — отрицательное отклонение. Условившись об этом, приступим к статистическим испытаниям. Каждое из испытаний состоит в одновременном подбрасывании трех монет. Предположим, что проделано N таких испытаний ($N\gg1$) и при этом в n испытаниях выпал герб одновременно у трех монет. Легко сообразить, что отношение $(N-n)/N$ и есть приближенное значение искомой вероятности.

Разумеется, вместо монеты можно использовать любой другой генератор случайных чисел. Можно было бы, например, подбрасывать три кубика, условившись связывать какие-то (неважно, какие именно) три грани с «положительным отклонением», а остальные три грани — с «отрицательным отклонением».

Подчеркнем, что в обоих рассмотренных примерах случай выступал в роли не отрицательного, а положительного фактора, в роли инструмента, позволяющего получить нужную информацию. Образно говоря, здесь случай работает на нас, а не против нас.

В игру вступает таблица случайных чисел. На практике в простых ситуациях, подобных описанным выше, никто не прибегает к методу статистических испытаний. Его используют тогда, когда рассчитать искомую вероятность очень трудно или даже вообще невозможно. Здесь мы предвидим вопрос читателя: а не окажутся ли тогда статистические испытания чрезмерно сложными и громоздкими? В рассмотренных примерах мы бросали крупинки или

занимались подбрасыванием трех монет. Что потребуется от нас в непростых ситуациях? Может быть, при этом возникнут практически непреодолимые трудности?

В действительности совсем не обязательно ставить эксперимент со статистическими испытаниями. Вместо реальных испытаний (бросаний крупинок, подбрасывания монет и т. п.) достаточно воспользоваться *таблицей случайных чисел*. Покажем, как это делается, в двух предложенных ранее примерах.

Первый пример. Обратимся к фигуре, которая рассматривалась на рисунке 2.6. Проведем вдоль сторон квадрата координатные оси. Масштаб выберем таким, чтобы сторона квадрата равнялась единице (рис. 2.7). Вместо того чтобы бросать крупинки, обратимся к таблице слунайных чисел (см. рис. 1.6). Предварительно каждое из чисел этой таблицы поделим на 10 000, чтобы иметь набор случайных чисел, находящихся в пределах от 0 до 1. Условимся рассматривать числа в нечетных строках таблицы в качестве x-координаты, а стоящие непосредственно под ними — в качестве y-координаты случайных точек, которые будем наносить на рисунок, постепенно перемещаясь по таблице случайных чисел (например, сначала вдоль всего первого столбца сверху вниз, затем вдоль всего второго столбца и т. д.). Первые пятнадцать случайных точек показаны на рисунке красным цветом, они имеют координаты: (0,0655; 0,5255), (0,6314; 0,3157), (0,9052; 0,4105), (0,1437; 0,4064), (0,1037; 0,5718), (0,5127; 0,9401), (0,4064; 0,5458), (0,2461; 0,4320), (0,3466; 0,9313), (0,5179; 0,3010), (0,9599; 0,4242), (0,3585; 0,5950), (0,8462; 0,0456), (0,0672; 0,5163), (0,4995; 0,6751). Черным цветом на рисунке показаны еще 85 случайных точек. Используя рисунок, нетрудно подсчитать, что для первых пятнадцати точек $N_2/N_1 = 13/15$ и, следовательно, $\pi = 3{,}47$; для ста точек $N_2/N_1 = 78/100$ и, следовательно, $\pi = 3{,}12$.

Второй пример. Вместо того чтобы подбрасывать монеты, воспользуемся уже знакомой читателю таблицей случайных чисел (см. рис. 1.6). Каждое из чисел таблицы, которое больше 5000, заменим знаком «+»; остальные числа заменим знаком «—». В результате мы получим таблицу, состоящую из случайного набора плюсов и минусов. Разобьем этот набор на тройки знаков так, как это показано на рисунке 2.8. Каждой тройке отвечает набор из трех деталей. Знак «+» означает, что соответствующая деталь с положительным отклонением, а знак «—» — с отрицательным. Приближенное значение искомой вероятности равно отношению $(N - n)/N$, где N — полное число троек, а n — число троек с тремя плюсами (на рисунке они закрашены). Видно, что в данном случае $(N - n)/N = 0{,}9$, что достаточно близко к точному значению 0,875.

Итак, мы свели статистические испытания к работе над таблицей случайных чисел, заменили экспериментальный стенд письменным столом. Вместо того чтобы реально проигрывать множество испытаний, мы просто читаем таблицу случайных чисел.

В игру вступает ЭВМ. Вместо того чтобы самим трудиться над

таблицей случайных чисел, можно поручить эту работу *вычислительной машине.* Введем таблицу случайных чисел в ЭВМ и предложим ей самой «просмотреть» и соответствующим образом. отсортировать случайные числа. Применительно к нашим двум примерам это будет выглядеть так.

Первый пример. ЭВМ должна проверить координаты x и y каждой случайной точки на предмет выполнения неравенства $x^2+y^2<1$. Она должна подсчитать число точек, для которых это неравенство выполняется (это число есть N_2), и число точек, для которых указанное неравенство не выполняется (число таких точек равно разности N_1-N_2).

Второй пример. Все заложенные в ЭВМ случайные числа должны быть разбиты на тройки. ЭВМ перебирает все эти тройки и выделяет те из них, где все три числа больше 5000. Число таких троек есть n.

Метод Монте-Карло. После того как в игру вступила ЭВМ, вся ситуация существенно изменяется. Работая с таблицей случайных чисел в соответствии с определенной программой, ЭВМ как бы проигрывает необходимые статистические испытания, делая это во много раз быстрее, чем это могло бы быть сделано на экспериментальном стенде или при обработке таблицы случайных чисел вручную. Вот теперь можно говорить о методе Монте-Карло — очень полезном и эффективном методе вероятностных расчетов, применяемом к самым различным практическим задачам и прежде всего к тем, которые нельзя решить аналитически.

Говоря о методе Монте-Карло, подчеркнем два обстоятельства. *Во-первых*, в данном методе мы используем *случайность против случайности.* Мы не пытаемся проникнуть в глубь сложных случайных процессов, не стараемся как-то смоделировать эти процессы. Вместо этого мы как бы предлагаем самой же случайности «разобраться» в тех сложностях, которые она породила. Случайность усложняет рассматриваемую картину, случайность же используется как инструмент исследования такой картины. *Во-вторых*, данный метод *универсален*, поскольку он не ограничен рамками каких-либо предположений, упрощений, моделей. Отсюда две основные области его применения. Первая область — исследование тех случайных процессов, которые в силу своей сложности не поддаются аналитическому рассмотрению. Вторая область — проверка правильности, степени точности аналитических моделей, применяемых в тех или иных конкретных ситуациях.

Метод Монте-Карло широко применяется при исследовании операций, при отыскании оптимальных решений в условиях неопределенности, при рассмотрении сложных многокритериальных задач. Этот метод с успехом используется также в современной физике — при исследовании сложных процессов, насыщенных случайностями.

Пример моделирования физического процесса по методу Монте-Карло. Рассмотрим распространение потока нейтронов через стенку ядерного реактора. В активной зоне реактора происходят акты деления ядер урана, сопровождающиеся рождением нейтронов с весь-

Рис. 2.8

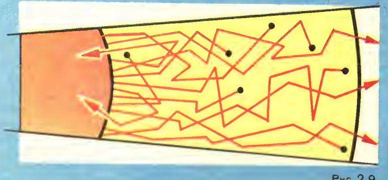


Рис. 2.9

ма высокими энергиями (порядка нескольких миллионов электрон-вольт). Реактор окружен стенкой, защищающей рабочее помещение от активной зоны. Стенка бомбардируется интенсивным нейтронным потоком, распространяющимся из активной зоны. Нейтроны, проникая внутрь стенки, взаимодействуют с ядрами атомов вещества стенки, в результате чего они могут быть либо поглощены, либо рассеяны. В последнем случае они передают часть своей энергии ядрам, на которых происходит рассеяние.

Перед нами сложный множественный физический процесс, *насыщенный случайностями*. Случайны энергия и направление движения нейтрона в момент его перехода из активной зоны в стенку, случайна длина пути нейтрона до первого взаимодействия, случаен характер взаимодействия (поглощение или рассеяние), случайны энергия и направление движения рассеянного нейтрона и т. д. Поясним в общих чертах, как применяют метод Монте-Карло к анализу таких процессов. Этот анализ основан на использовании ЭВМ, куда предварительно вводят сведения об элементарных процессах взаимодействия нейтронов с ядрами вещества (вероятности поглощения, вероятности рассеяния и т. п.), а также параметры падающего на стенку нейтронного потока и параметры вещества стенки. Для некоторого нейтрона разыгрываются, т. е. случайно выбираются с учетом соответствующих вероятностей, исходные энергия и направление движения (в момент перехода из активной зоны реактора в стенку). Затем разыгрывается (опять-таки с учетом соответствующих вероятностей) длина пробега нейтрона до первого взаимодействия. Вслед за этим разыгрывается тип первого взаимодействия. Если при этом нейтрон не поглотился, то разыгрываются все последующие события: длина пробега нейтрона до второго взаимодействия, тип второго взаимодействия и т. д. В результате определяется «судьба» рассматриваемого нейтрона от момента его попадания внутрь стенки до момента выхода из игры. Имеются три возможных выхода из игры — поглощение, рассеяние назад в активную зону реактора и рассеяние в рабочее помещение. Затем вся эта процедура машинных розыгрышей повторяется для второго, третьего, четвертого и многих других нейтронов. В итоге мы получаем множество возможных траекторий нейтронов внутри стенки (рис. 2.9). Каждая из траекторий — результат одного статистического испытания, представляющего собой розыгрыш судьбы отдельного нейтрона. Огромный набор таких испытаний позволяет проанализировать в целом картину прохождения нейтронного потока через защитную стенку и получить, в частности, рекомендации по толщине стенки и составу ее вещества, гарантирующие безопасность работы в помещении, где находится реактор.

Современная физика дает нам много примеров применения метода Монте-Карло. Физики обращаются к этому методу, когда изучают развитие ливней частиц космического излучения в земной атмосфере, исследуют поведение больших потоков электронов в современных электровакуумных приборах, имеют дело с различными цепными (саморазвивающимися) реакциями.

Игра и принятие решения

Что такое «теория игр»? Предположим, что надо принимать решение в условиях, когда нашим целям противостоят противоположные цели другой стороны, когда нашей воле противостоит другая воля. Подобные ситуации встречаются часто, это — *конфликтные ситуации.* Они характерны для военных действий, игровых видов спорта, практической деятельности. С той или иной степенью остроты они часто возникают при решении различных экономических и политических проблем.

Хоккеист на игровой площадке принимает то или иное решение, исходя не только из сложившейся к данному моменту обстановки, но и из возможных действий со стороны других игроков. Шахматист, принимая всякий раз то или иное решение, старается учесть возможные действия противника. Решение о той или иной военной операции должно приниматься с учетом ответных действий противной стороны. Принимая решение об установлении цены товара, продавец на рынке учитывает возможную реакцию покупателя на эту цену. Проводя те или иные мероприятия в предвыборной кампании, любая политическая партия в капиталистической стране старается предвидеть действия других партий, участвующих в борьбе за власть. Во всех подобных случаях происходит столкновение противоположных интересов; принятие решения каждой из сторон связано с *преодолением конфликта.*

Принятие решения в конфликтной ситуации затрудняется из-за *неопределенности поведения противника.* Мы знаем, что противник стремится предпринять наименее выгодные для нас действия, с тем чтобы обеспечить себе наибольший успех. Но мы не знаем, в какой мере наш противник умеет оценивать обстановку и возможные последствия и, в частности, как он оценивает наши намерения и возможности. Мы не можем точно предсказать действия противника, равно как и он не может точно предсказать наши действия. И тем не менее как нам, так и ему приходится принимать вполне определенные решения.

Необходимость обоснования *оптимальных* решений, принимаемых в тех или иных конфликтных ситуациях, привела к возникновению специального направления в современной математике — *теории игр.* Под термином «игра» здесь понимается упрощенная математическая модель рассматриваемой конфликтной ситуации. В отличие от реального конфликта игра ведется по определенным правилам, которые четко определяют права и обязанности участников игры, а также исход игры (выигрыш и проигрыш каждого участника). Задолго до появления теории игр широко использовались подобные упрощенные модели конфликтов — йгры в буквальном смысле слова: шахматы, шашки, домино, карточные игры и т. д. Собственно говоря, отсюда и происходит как название самой теории, так и различные термины, используемые в ней. Так, конфликтующие стороны называют «игроками», одну реализацию игры «партией», выбор игроком того или иного действия (в пределах правил) «ходом» и т. д.

Различают два вида ходов — личные и случайные. *Личный* ход предполагает сознательный выбор игроком того или иного действия, разрешенного правилами игры. *Случайный* ход не зависит от воли игрока — он может быть определен по результату бросания монеты или игральной кости, вынимания карты из колоды и т. п. Игры, состоящие только из случайных ходов, называют *азартными*. Характерный пример — игра в лото. Игры, в которых имеются личные ходы, называются *стратегическими*. Существуют стратегические игры, состоящие только из личных ходов (например, шахматы). Существуют также стратегические игры, состоящие как из личных, так и из случайных ходов (например, карточные игры). Заметим, что в играх с личными и случайными ходами неопределенность выступает как бы в двух обличьях — в виде неопределенности результата случайных ходов и в виде неопределенности поведения противника в его личных ходах.

Теория игр не интересуется азартными играми. Она занимается только стратегическими играми. Задача теории игр — определить такую стратегию игрока, при которой его шансы на выигрыш оказались бы наибольшими. В основе поиска оптимальных стратегий лежит следующее основное положение. Считается, что противник так же разумен и активен, как и сам игрок, и предпринимает все меры для того, чтобы достичь успеха.

Разумеется, на практике это не всегда выполняется. Часто наши действия в реальном конфликте оказываются оптимальными не тогда, когда мы исходим из наиболее разумного поведения противника, а тогда, когда нам удается угадать, в чем противник оказывается «глупым», и удается воспользоваться этой «глупостью». При этом мы рискуем. Известно, как рискованно рассчитывать на глупость противника. Теория игр не учитывает элементов риска. Она выявляет лишь наиболее осторожные, «перестраховочные» варианты поведения в данной ситуации. Можно сказать, что теория игр дает нам мудрые советы. Учитывая эти советы, мы затем принимаем на практике те или иные решения, часто идя сознательно на некоторый риск. Как пишет Е. С. Вентцель в своей книге «Исследование операций»: «Теория игр ценна прежде всего самой постановкой задач, которая учит не забывать о том, что противник тоже мыслит, и учитывать его возможные хитрости и уловки. Пусть рекомендации, вытекающие из игрового подхода, не всегда определенны и не всегда осуществимы — все же полезно, выбирая решение, ориентироваться, в числе других, и на игровую модель. Не надо только выводы, вытекающие из этой модели, считать окончательными и непререкаемыми».

Платежная матрица игры. В теории игр наиболее исследованы *конечные пáрные игры с нулевой суммой*. Игра называется *пáрной*, если в ней участвуют два игрока. Игра называется *конечной*, если у каждого игрока есть конечное число стратегий, т. е. конечное число вариантов поведения. Делая личный ход, игрок следует одной из стратегий. Игра с *нулевой суммой* есть игра, в которой выигрыш одного игрока равен проигрышу другого.

Предположим, что рассматривается некоторая конечная па́рная игра с нулевой суммой, где у игрока A имеется m стратегий, а у игрока B — n стратегий (игра $m \times n$). Обозначим стратегии игрока A через A_1, A_2, ..., A_m, а стратегии игрока B через B_1, B_2, ..., B_n. Пусть игрок A, делая личный ход, выбирает некоторую стратегию A_i $(1 \leqslant i \leqslant m)$, а игрок B выбирает при этом некоторую стратегию B_j $(1 \leqslant j \leqslant n)$. Обозначим через a_{ij} реализуемый в этом случае выигрыш игрока A. Для определенности мы будем отождествлять себя с игроком A и рассматривать каждый ход с позиции выигрыша игрока A. При этом под выигрышем a_{ij} может пониматься как действительный выигрыш, так и проигрыш (например, проигрыш может выступать как отрицательный выигрыш). Набор выигрышей a_{ij} для разных значений i и j располагают в виде матрицы, строки которой отвечают стратегиям игрока A, а столбцы — стратегиям игрока B (рис. 2.10). Это есть *платежная матрица* игры.

В качестве примера рассмотрим следующую игру. Игроки A и B одновременно и независимо друг от друга записывают каждый одно из трех чисел: либо 1, либо 2, либо 3. Если сумма записанных чисел оказывается *четной*, то игрок B платит игроку A эту сумму; если же сумма чисел оказывается *нечетной*, то эту сумму выплачивает игрок A игроку B. У игрока A три стратегии: A_1 — записать 1, A_2 — записать 2, A_3 — записать 3. Стратегии игрока B аналогичны. Рассматриваемая игра есть игра 3×3, ее платежная матрица имеет 3 строки и 3 столбца. Эта матрица представлена на рисунке 2.11, *а*. Заметим, что выигрыш игрока A, равный, например, -3, означает в действительности его проигрыш, так как в этом случае игрок A выплачивает 3 копейки игроку B.

В матрице на рисунке 2.11, *а* одни элементы являются положительными, а другие отрицательными. Можно сделать так, чтобы все элементы платежной матрицы были положительными. Для этого увеличим каждый элемент рассматриваемой матрицы на одно и то же число, например на 6. Получим матрицу, представленную на рисунке 2.11, *б*. С точки зрения анализа оптимальных стратегий эта матрица эквивалентна исходной.

Принцип минимакса. Будем анализировать игру, используя платежную матрицу, показанную на рисунке 2.11, *б*. Предположим, что мы (игрок A) выбираем стратегию A_1. Тогда в зависимости от того, какую стратегию изберет противник, наш выигрыш будет равен либо 8, либо 3, либо 10. Итак, выбирая стратегию A_1, мы в худшем случае получаем выигрыш 3. Если же мы выберем стратегию A_2 или стратегию A_3, то будем иметь в худшем случае выигрыш 1. Запишем минимальные возможные выигрыши для разных стратегий A_i в виде дополнительного столбца платежной матрицы (рис. 2.12). Ясно, что следует выбирать ту стратегию, где *минимальный возможный выигрыш оказывается наибольшим* (по сравнению с остальными стратегиями). В данном случае это есть стратегия A_1. Выигрыш 3 является максимальным в тройке минимальных выигрышей (в тройке 3, 1, 1). Его называют *макси-*

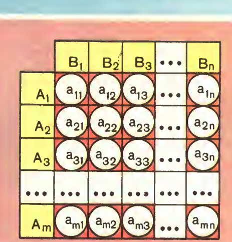


Рис. 2.10

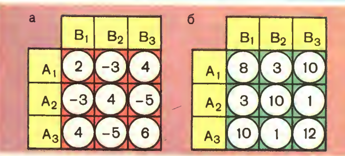


Рис. 2.11

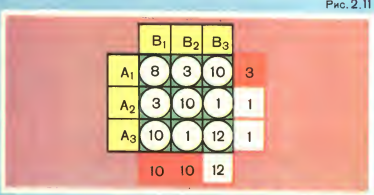


Рис. 2.12

минным выигрышем или, проще, *максимином*. Есть у него еще одно название — *нижняя цена игры*. Итак, если мы выбираем максиминную стратегию (в данном случае это есть стратегия A_1), то при любом поведении противника нам гарантирован выигрыш не меньше нижней цены игры (в данном случае этот выигрыш равен 3). Аналогичным образом рассуждает противник. Если он выберет стратегию B_1, то в худшем для себя случае позволит нам получить выигрыш 10. То же можно сказать и о стратегии B_2. При выборе стратегии B_3 худший (для противника) случай соответствует нашему выигрышу, равному 12. Числа 10, 10, 12 — максимальные значения наших выигрышей, отвечающие стратегиям противника B_1, B_2, B_3 соответственно. Выпишем эти значения в виде дополнительной строки платежной матрицы (см. рис. 2.12). Ясно, что противник должен выбрать ту стратегию, где наш *максимальный возможный выигрыш оказывается наименьшим*. Это есть либо стратегия B_1, либо B_2. Обе стратегии являются минимаксными, обе они дают противнику гарантию, что наш выигрыш в любом случае не превысит *минимакса*, или, иначе, *верхней цены игры*, равной в данном случае 10.

Наша максиминная стратегия, равно как и минимаксная стратегия противника, является наиболее осторожной, «перестраховочной» стратегией. Принцип осторожности, диктующий игрокам выбор таких стратегий, называют *принципом минимакса*.

Но вернемся к матрице на рисунке 2.12 и попробуем немного порассуждать. У противника две минимаксных стратегии — B_1 и B_2. Какую он скорее всего выберет? Полагая, что мы проявили осторожность и выбрали максиминную стратегию A_1, он, возможно, не станет выбирать стратегию B_1, поскольку при этом мы получили бы выигрыш 8. Значит, скорее всего он выберет стратегию B_2, тогда наш выигрыш будет равен 3. Но если мы правильно поняли замыслы противника, то не следует ли нам рискнуть и выбрать стратегию A_2? Ведь при выборе противником стратегии B_2 наша стратегия A_2 позволит нам получить выигрыш 10. Однако наше отступление от принципа минимакса может дорого нам обойтись. Если противник окажется достаточно хитроумным и проделает такие же рассуждения, то он может ответить на нашу стратегию A_2 не стратегией B_2, а стратегией B_3. И тогда вместо выигрыша 10 мы получим выигрыш всего лишь 1.

Означает ли все это, что теория игр рекомендует придерживаться только минимаксных (максиминных) стратегий? Ответ на этот вопрос зависит от того, имеет или не имеет платежная матрица игры *седловую точку*.

Игра с седловой точкой. Рассмотрим некоторую игру 3×3, платежная матрица которой дана на рисунке 2.13. Здесь как максиминный, так и минимаксный выигрыши равны 4. Иными словами, в данной игре нижняя и верхняя цена игры совпадают, обе равны 4. Выигрыш 4 является одновременно и максимальным из минимальных выигрышей для стратегий A_1, A_2, A_3, и минимальным из максимальных выигрышей для стратегий B_1, B_2, B_3. В геомет-

рии точку на поверхности, являющуюся одновременно минимумом по одной оси координат и максимумом по другой, называют *седловой точкой*. Такой точкой является точка C на поверхности, изображенной на том же рисунке. Точка C есть максимум по x-координате и минимум по y-координате. Легко видеть, что поверхность в окрестности этой точки действительно похожа на седло. По аналогии с геометрией элемент $a_{22}=4$ рассматриваемой здесь платежной матрицы называют *седловой точкой матрицы*, а об игре говорят, что она имеет седловую точку.

Достаточно посмотреть внимательно на матрицу, показанную на рисунке 2.13, чтобы понять, что каждый из игроков должен придерживаться максиминной (минимаксной) стратегии. Эти стратегии являются оптимальными в игре с седловой точкой. Любое отклонение от них будет невыгодно для игрока, допустившего отклонение.

Если же игра не имеет седловой точки (см. матрицу на рис. 2.12), то ни одна из стратегий A_i или B_j не является оптимальной. **Необходимость случайного изменения стратегии в игре без седловой точки.** Допустим, что мы и наш противник многократно играем в игру, матрица которой дана на рисунке 2.12. Если мы выберем определенную стратегию, например максиминную стратегию A_1, и будем придерживаться ее от игры к игре, то противник, поняв это, будет выбирать каждый раз стратегию B_2, в результате чего наш выигрыш не превысит нижней цены игры, т. е. будет равен 3. Если, однако, мы внезапно (для противника) сменим стратегию A_1 на стратегию A_2, то получим выигрыш 10. Разгадав нашу новую стратегию (если, конечно, мы начнем ее придерживаться в дальнейшем), противник тут же сменит стратегию B_2 на стратегию B_3, уменьшив наш выигрыш до 1. И так далее. Здесь проявляется общее правило для игр без седловой точки: игрок, играющий по *определенной (детерминированной) стратегии*, оказывается в более худшем положении по сравнению с игроком, который *меняет стратегию случайным образом*.

Впрочем, случайные изменения стратегии надо делать не как попало, а с умом. Пусть A_1, A_2, ..., A_m— возможные стратегии игрока A (см. рис. 2.10). Для получения наибольшего эффекта он должен использовать все или некоторые из этих стратегий случайным образом, но не с одинаковыми, а с разными (специально вычисленными) вероятностями. Пусть стратегия A_1 используется с вероятностью p_1, стратегия A_2 с вероятностью p_2 и т. д. Говорят, что игрок A применяет *смешанную стратегию* $S_A(p_1, p_2, ..., p_m)$. В отличие от смешанных стратегий S_A стратегии A_j называют *чистыми*. При надлежащем подборе вероятностей p_j смешанная стратегия может оказаться *оптимальной*. При этом выигрыш игрока A будет не меньше некоторого значения ν, называемого *ценой игры*. Это значение больше нижней цены игры, но меньше верхней.

Аналогичным образом должен вести себя игрок B. Его оптимальная стратегия также есть некоторая смешанная стратегия. Обоз-

начим ее как S_B $(q_1, q_2, ..., q_n)$, где q_j — специально подобранные вероятности, с которыми игрок B использует стратегии B_j При выборе игроком B оптимальной смешанной стратегии выигрыш игрока A будет не больше цены игры ν.

Поиск оптимальной смешанной стратегии. Обозначим через $S_A(p_1, ..., p_m)$ оптимальную смешанную стратегию игрока A. Требуется найти вероятности $p_1, p_2, ..., p_m$ и определить цену игры ν при условии, что известна платежная матрица игры (см. рис. 2.10). Допустим, что игрок B выбирает чистую стратегию B_1. Тогда средний выигрыш игрока A будет равен $a_{11}p_1+a_{21}p_2+...+a_{m1}p_m$. Этот выигрыш должен быть не меньше цены игры ν, следовательно,

$$a_{11}p_1+a_{21}p_2+...+a_{m1}p_m \geqslant \nu.$$

Если игрок B выберет стратегию B_2, то и в этом случае средний выигрыш игрока A должен быть не меньше цены игры ν, следовательно,

$$a_{12}p_1+a_{22}p_2+...+a_{m2}p_m \geqslant \nu.$$

Какую бы стратегию ни выбирал игрок B, выигрыш игрока A всегда должен быть не меньше цены игры ν. Поэтому мы можем записать следующую систему из n неравенств (напоминаем, что n — число чистых стратегий игрока B):

$$\left.\begin{array}{l} a_{11}p_1+a_{21}p_2+...+a_{m1}p_m\geqslant \nu; \\ a_{12}p_1+a_{22}p_2+...+a_{m2}p_m\geqslant \nu; \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ a_{1n}p_1+a_{2n}p_2+...+a_{mn}p_m\geqslant \nu. \end{array}\right\} \qquad (2.10)$$

При этом

$$p_1+p_2+...+p_m=1. \qquad (2.11)$$

Введя обозначения $x_1=p_1/\nu$, $x_2=p_2/\nu, ..., x_m=p_m/\nu$, перепишем (2.10) и (2.11) в виде:

$$\left.\begin{array}{l} a_{11}x_1+a_{21}x_2+...+a_{m1}x_m\geqslant 1; \\ a_{12}x_1+a_{22}x_2+...+a_{m2}x_m\geqslant 1; \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ a_{1n}x_1+a_{2n}x_2+...+a_{mn}x_m\geqslant 1; \end{array}\right\} \qquad (2.12)$$

$$x_1+x_2+...+x_m=1/\nu. \qquad (2.13)$$

Нам желательно, чтобы цена игры ν была как можно больше, следовательно, $1/\nu$ должна быть как можно меньше. Таким образом, поиск оптимальной смешанной стратегии свелся к решению следующей математической задачи: надо найти неотрицательные величины $x_1, x_2, ..., x_m$ такие, чтобы они удовлетворяли неравенствам (2.12) и при этом обращали в минимум сумму $x_1+x_2+...+x_m$.

Самолеты против зениток. Найдем оптимальную смешанную стратегию для некоторой конкретной игры. Предположим, что сторона A нападает на сторону B. У стороны A имеются два самолета,

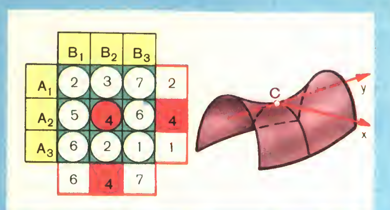


Рис. 2.13

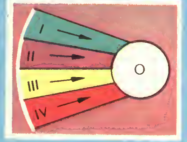


Рис. 2.14

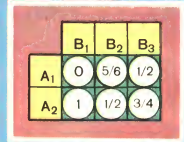


Рис. 2.15

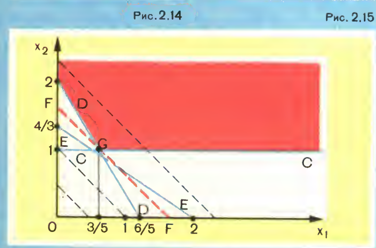


Рис. 2.16

несущие мощное поражающее средство. У стороны B имеются четыре зенитки, при помощи которых осуществляется оборона важного объекта. Чтобы объект оказался разрушенным, достаточно, чтобы к нему прорвался хотя бы один самолет. Для подхода к объекту самолеты могут выбрать любой из четырех воздушных коридоров (рис. 2.14; O — объект, I, II, III, IV — воздушные коридоры). Сторона A может послать оба самолета в одном и том же коридоре или направить их по разным коридорам. Сторона B может разместить свои четыре зенитки в пределах рассматриваемых коридоров разными способами. Каждая зенитка может произвести только один выстрел. Этот выстрел с достоверностью поражает самолет, если тот оказался в данном коридоре.

У стороны A есть две чистые стратегии: стратегия A_1 — самолеты посылаются в разных воздушных коридорах (безразлично, каких именно), стратегия A_2 — оба самолета посылаются в каком-то одном из коридоров. Возможные стратегии стороны B таковы: B_1 — поставить по зенитке на каждый коридор, B_2 — поставить по две зенитки на какие-то два коридора (остальные два коридора остаются неохраняемыми), B_3 — поставить две зенитки на один из коридоров и по одной зенитке еще на два коридора, B_4 — поставить три зенитки на один из коридоров и одну зенитку еще на один коридор, B_5 — поставить все четыре зенитки на один из коридоров. Стратегии B_4 и B_5 заведомо невыгодны хотя бы потому, что три, а тем более четыре зенитки в пределах одного коридора не нужны, ведь у стороны A всего два самолета. Поэтому ограничимся стратегиями B_1, B_2, B_3.

Предположим, что сторона A выбрала стратегию A_1, а сторона B — стратегию B_1. Ясно, что тогда ни один самолет не прорвется к объекту — выигрыш стороны A есть нуль ($a_{11}=0$). Пусть выбраны стратегии A_1 и B_2. Допустим при этом, что зенитки находятся в коридорах I и II. Самолеты летят в разных коридорах, причем равновероятны шесть вариантов: они летят в коридорах I и II, летят в коридорах I и III, в коридорах I и IV, в II и III, в II и IV, в III и IV. Только в одном из указанных шести случаев ни один из самолетов не прорвется к объекту (когда они летят в коридорах I и II). Какие бы два коридора ни выбирала сторона B для размещения пар зениток, всегда у самолетов будут шесть равновероятных вариантов и только один из них проигрышный. Таким образом, при выборе стратегий A_1 и B_2 вероятный выигрыш стороны A составляет 5/6 ($a_{12}=5/6$). Рассуждая подобным образом, нетрудно найти остальные элементы платежной матрицы данной игры. Матрица показана на рисунке 2.15, это есть матрица 2×3. Заметим, что элементы матрицы — *вероятностные* выигрыши; здесь уже чистые стратегии включают в себя случайность. Нижняя цена игры равна 1/2, верхняя равна 3/4. Максиминная стратегия есть A_2, минимаксная — B_3. Седловой точки нет, оптимальное решение игры лежит в области смешанных стратегий.

Чтобы найти оптимальную смешанную стратегию, воспользуемся видом платежной матрицы и соотношениями (2.12) и (2.13). В

данном случае эти соотношения принимают вид:

$$x_2 \geqslant 1; \quad \frac{5}{6}x_1 + \frac{1}{2}x_2 \geqslant 1; \quad \frac{1}{2}x_1 + \frac{3}{4}x_2 \geqslant 1; \tag{2.14}$$

$$x_1 + x_2 = 1/\nu. \tag{2.15}$$

Решение удобно представить графически. Будем откладывать положительные значения x_1 и x_2 на осях координат (рис. 2.16). Первому неравенству (2.14) отвечает область выше прямой CC; второму неравенству (2.14) — выше прямой DD; третьему неравенству (2.14) — выше прямой EE. Все три неравенства выполняются в пределах области, закрашенной на рисунке красным цветом. Уравнение $x_1+x_2=\text{const}$ описывает семейство параллельных прямых, которые показаны на рисунке штриховыми линиями. Из всех таких прямых, имеющих хотя бы одну точку в пределах красной области, наименьшей сумме x_1+x_2 отвечает прямая FF. Точка G фиксирует решение, соответствующее *оптимальной смешанной стратегии*. Координаты этой точки: $x_1=3/5$, $x_2=1$. Отсюда находим: $\nu=5/8$. $p_1=3/8$, $p_2=5/8$. Итак, оптимальная смешанная стратегия стороны A предполагает использование стратегии A_1 с вероятностью 3/8 и стратегии A_2 с вероятностью 5/8.

Как воспользоваться этой рекомендацией на практике? Если игра происходит *один раз*, то стороне A следует, по-видимому, избрать стратегию A_2, ведь $p_2>p_1$. Предположим, что данная игра совершается *многократно* (например, по отношению к многим объектам, подлежащим бомбардировке). Если игра повторяется N раз ($N\gg 1$), то в $3N/8$ случаях сторона A должна избрать стратегию A_1, а в $5N/8$ случаях стратегию A_2.

До сих пор мы обсуждали поведение лишь стороны A, предоставляя стороне B действовать произвольно. При выборе стороной A оптимальной смешанной стратегии ее средний выигрыш оказывается в пределах между верхней ценой игры, равной 3/4, и ценой игры $\nu=5/8$. При неразумном поведении стороны B выигрыш стороны A может оказаться равным верхней цене игры (и даже может стать больше). Если же сторона B, в свою очередь, будет придерживаться оптимальной смешанной стратегии, то выигрыш стороны A окажется равным цене игры ν. Оптимальная смешанная стратегия стороны B сводится к тому, что эта сторона вообще не применяет стратегию B_3, стратегию B_1 использует с вероятностью 1/4, а стратегию B_2 с вероятностью 3/4. Нецелесообразность применения стратегии B_3 усматривается из рисунка 2.16: соответствующая этой стратегии прямая EE не имеет общих точек с красной областью. Для определения вероятностей, с какими должны использоваться стратегии B_1 и B_2, воспользуемся уже найденным значением цены игры $(\nu=5/8): q_1 \cdot 0 + (1-q_1)\frac{5}{6} = \frac{5}{8}$. Отсюда видно, что $q_1 = 1/4$, $q_2 = 1 - q_1 = 3/4$.

ГЛАВА 3 Управление и самоуправление

Процесс получения и использования информации является процессом нашего приспособления к случайностям внешней среды и нашей жизнедеятельности в этой среде.

Н. Винер

Кибернетика проникла и продолжает проникать во все области трудовой деятельности и жизни человека. Это наука об оптимальном управлении сложными процессами и системами.

А. И. Берг

Проблема управления

Управление против дезорганизации. Хотя окружающий нас мир и насыщен случайностями, тем не менее он оказывается достаточно организованным, во многих отношениях упорядоченным. *Дезорганизующему действию случайностей противостоит организующее действие процессов управления и самоуправления.*

Самолет летит, скажем, из Москвы в Ленинград. В полете на него действуют разные факторы случайного характера. Поэтому все три пространственные координаты самолета оказываются случайными функциями времени. Траектория данного полета есть одна из реализаций этих случайных функций. Все эти «тонкости» нисколько не волнуют пассажиров. Застегивая ремни перед взлетом, они не сомневаются, что, какие бы грозовые явления ни встретились на трассе полета, какие бы ветры ни сбивали самолет с курса, все равно он прибудет в ленинградский аэропорт. Основание для такой уверенности — система управления самолетом, действия пилотов. Ранее мы познакомились с системами массового обслуживания. Казалось бы, эти системы насыщены случайностями. Тем не менее они выполняют свои задачи. Это объясняется продуманной организацией систем, управлением их работой.

Фактор управления может выступать в разных обличьях. Мы хотим, чтобы книга дольше служила людям. Этому препятствуют различного рода случайности как чисто физической природы, так и связанные с отношением читателей к книгам. И вот мы начинаем управлять: заботимся о переплете, регулируем должным образом температуру, влажность, освещенность в помещении, где хранятся книги, заводим на книгу формуляр, устанавливаем правила пользования книгами.

Нет такого человека, который был бы гарантирован от заболеваний. И хотя многие заболевания имеют вполне определенные причины, картина заболеваний в масштабах, скажем, города характеризуется обилием случайностей. Вступая в борьбу с ними, мы начинаем управлять: заботимся об улучшении условий жизни и труда людей, проводим медицинские профилактические мероприя-

тия, строим стадионы, бассейны, спортивные комплексы, обеспечиваем аптеки необходимыми лекарствами.

Итак, в мире совершается *противоборство* двух мощных факторов, двух основных тенденций. С одной стороны, *фактор случайного*, тенденция к дезорганизации, разупорядочиванию и в конечном счете к разрушению. С другой стороны, *фактор управления* и самоуправления, тенденция к организованности, упорядочиванию, дальнейшему развитию, прогрессу.

Выбор как необходимое условие управления. Если бы все процессы и явления в мире были строго детерминированы, то было бы бессмысленно говорить о самой возможности управления. *Чтобы управлять, надо иметь выбор*. Нет смысла говорить о принятии того или иного решения, если все заранее предопределено. У каждого явления должна быть вероятность различных линий развития. Можно сказать, что мир, построенный на вероятности, и есть тот самый мир, в котором только и возможно управление.

Управление действует против случайностей. Вместе с тем возможность управления предопределяется самим существованием случайностей. Именно случайности помогают избежать предопределенности. Получается, что случайности как бы «порождают» своего «могильщика» — управление. Мы видим в этом проявление диалектического единства необходимого и случайного в реально существующем мире.

Проповедуя божественное происхождение всего сущего, связанную с «божьей волей» предопределенность всего происходящего, церковь пытается навязать людям представление о строго детерминированном мире, где у людей не остается свободного выбора, а следовательно, нет и самой возможности чем-либо управлять. Такая точка зрения лишает людей воли, желания действовать или, тем более, чему-либо противодействовать. Надо ли доказывать ложность и вредность подобной точки зрения?

Управление и обратная связь. На рисунке 3.1 даны в упрощенном виде две принципиально разные схемы управления: *С* — система, которой управляют; *УУ* — управляющее устройство; *V* — входное воздействие на управляемую систему (управляющее воздействие); *P* — случайные возмущения, воздействующие на управляемую систему; *W* — выходная величина, выступающая как результат управления системой. В отличие от схемы *а*, в схеме *б* имеется *обратная связь:* управляющее устройство получает сведения о результатах управления.

Зачем нужна обратная связь? Отвечая на этот вопрос, отметим, что «взаимоотношения» случайного и управления имеют характер активного противоборства. Управление активно действует против случайностей, случайности столь же активно действуют против управления. Последнее обстоятельство требует от управления гибкости, способности по ходу дела перестраиваться. Для перестройки надо, чтобы управляющее устройство все время получало сведения о результатах управления и в соответствии с этим корректировало свои воздействия на систему.

а

УУ
V
C
P
W

б

УУ
V
C
P
W
ОБРАТНАЯ СВЯЗЬ

Рис. 3.1

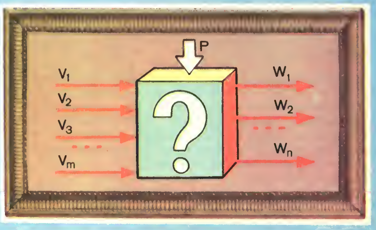


Рис. 3.2

По сути дела, любая реальная схема управления предполагает наличие обратной связи. *Управление без обратной связи не только неэффективно, но и фактически всегда нежизнеспособно.*

Вот простой пример: автомобиль, управляемый водителем. Представьте на минуту, что обратная связь вдруг исчезла — водитель перестал следить за тем, как движется управляемый им автомобиль. Управление автомобилем продолжается, но без обратной связи. Немедленно начинается атака со стороны всевозможных случайностей. Случайная выбоина на шоссе, случайный поворот, случайно появившийся встречный автомобиль — все эти случайности в считанные секунды приведут к аварии.

Алгоритм управления. Что и как надо делать, чтобы осуществить управление? Ответ на этот вопрос определяется конкретной ситуацией и той целью, которую мы преследуем в том или ином случае. Этот ответ содержится в алгоритме управления. *Алгоритм управления есть последовательность определенных действий, которые необходимо выполнить, чтобы достичь поставленной цели.*

В примере с водителем, управляющим автомобилем, алгоритм управления содержит в себе правила включения и выключения двигателя, торможения, осуществления поворотов, переключения скоростей и т. д. Этот алгоритм содержит в себе также правила уличного движения.

В одних случаях алгоритм управления довольно прост. Например, чтобы воспользоваться автоматом по продаже газированной воды, надо выполнить всего три действия: поставить в определенное место стакан, опустить в щель монету и нажать соответствующую кнопку. Этим и исчерпывается алгоритм управления данным устройством. В других случаях алгоритм управления значительно сложнее. Так, более сложно управлять автомобилем. Еще более сложно управлять реактивным самолетом. В особо сложных случаях алгоритм управления не может быть даже определен исчерпывающим образом. Не существует, например, полностью определенных алгоритмов управления крупным предприятием или, тем более, отраслью промышленности.

От «черного ящика» к кибернетике

Несмотря на разнообразие алгоритмов, процессы управления могут исследоваться с общих позиций, независимо от особенностей рассматриваемых систем. Характерный пример — моделирование различных систем на основе использования модели «черного ящика».

Что такое «черный ящик»? Предположим, что рассматривается некоторая управляемая система. Пусть V_1, V_2, ..., V_m — входные воздействия на систему (управляющие воздействия), P — случайные воздействия, W_1, W_2, ..., W_n — выходные величины системы (рис. 3.2). Предположим далее, что мы не знаем или попросту не желаем знать внутреннее устройство системы. Мы исследуем лишь связи между входными воздействиями (V_1, V_2...)

и выходными величинами (W_1, W_2...). В этом случае говорят, что данная система уподоблена «черному ящику».

Под «черным ящиком» понимают любую управляемую систему, если ее внутреннее устройство не рассматривается, а исследуются только *реакции выходных величин на выходные воздействия.*

Человек в окружении «черных ящиков». Благодаря развитию науки и техники современный человек оказался в окружении огромного количества разнообразных управляемых систем. Как правило, это его нисколько не тяготит, поскольку он быстро привыкает (подчас даже не сознавая того) рассматривать все эти системы как «черные ящики». Он знает, как, что, где надо повернуть, нажать, переключить, чтобы получить нужный эффект. Чтобы посмотреть телепередачу, нет нужды знать устройство и принцип работы телевизора. Достаточно нажать соответствующую кнопку и повернуть переключатель программ. Чтобы поговорить по телефону, не надо предварительно изучать его устройство. Надо снять телефонную трубку и, дождавшись гудка, набрать нужную последовательность цифр на телефонном диске. И телевизор, и телефон, и многие другие управляемые системы выступают перед нами в роли «черных ящиков». Разумеется, при желании можно разобраться в устройстве систем, понять принцип их работы. Но современному человеку часто жаль времени на изучение того, без чего он может вполне обойтись в своей практической деятельности. Он все чаще предпочитает пользоваться «черными ящиками», а в случае их поломки обращается к соответствующим специалистам.

Следует признать некоторую справедливость сетований на то, что современный человек становится менее любознательным, что он не очень стремится проникнуть в глубь вещей, поскольку «вещей» этих накапливается слишком много, а пользоваться ими вовсе несложно. Однако не нужно чрезмерно сгущать краски. Во-первых, существует система всеобщего среднего образования, обеспечивающая каждого человека минимумом фундаментальных знаний. Во-вторых, с точки зрения развития общества важно не то, что знает тот или иной человек, а то, какими знаниями располагает общество в целом.

Сложные системы и «черный ящик». Современные системы оказываются все более и более сложными, а их функциональные возможности (то, что они могут делать) становятся все более богатыми. Естественно, что в этих условиях на первый план выдвигается исследование именно функциональных возможностей систем. Исследование же внутренней структуры систем отступает на второй план, тем более что во многих случаях такое исследование в полном объеме оказывается практически невозможным из-за сложности систем.

Подобное смещение акцентов приводит нас к качественно новой точке зрения, когда главной задачей становится изучение общих закономерностей процессов управления и самоуправления независимо от конкретного устройства тех или иных управляемых систем.

Именно эта точка зрения и приводит к *кибернетике* как науке об управлении (самоуправлении) в сложных системах.

При этом выявляется одно весьма любопытное обстоятельство, заставляющее нас по-иному взглянуть на модель «черного ящика». Оказывается, что вовсе не обязательно вникать во все тонкости структуры достаточно сложной системы, что расчленение ее на составные части может привести к утрате принципиально важной информации. В этой ситуации модель «черного ящика» приобретает *принципиальное* значение — как единственно приемлемый подход к анализу сложной системы.

Что такое кибернетика? Создание *кибернетики* связывают с именем видного американского ученого Норберта Винера (1894—1964). Принято считать, что история кибернетики начинается с 1948 года, когда Винер опубликовал свою знаменитую книгу под названием «Кибернетика, или управление и связь в животном и машине». Как писал Винер, «было решено назвать всю теорию управления и связи в машинах и живых организмах кибернетикой», что в переводе с греческого — «кормчий».

Следует отметить, что сам по себе термин «кибернетика» не нов. Он встречается уже у Платона, где обозначает искусство управления кораблем. В первой половине XIX века французский физик Ампер, занимаясь классификацией наук, поместил в своей системе в рубрике за номером 83 науку, которая должна исследовать способы управления государством. Ампер назвал эту науку кибернетикой. Сегодня мы используем термин «кибернетика» только в том смысле, какой был дан ему Винером. *Кибернетика есть наука об общих закономерностях процессов управления и связи в сложных системах, включая как машины, так и живые организмы.*

В книге «Этот случайный, случайный, случайный мир», написанной советским ученым Л. А. Растригиным, есть очень точное замечание: «До появления кибернетики процессы управления в электрическом генераторе рассматривались электротехникой, управление движением часового маятника исследовалось в механике, управление динамикой популяций — в биологии. Винер впервые указал на универсальность управления и показал, что процесс упорядочения объекта можно производить стандартными приемами, т. е. применять методы кибернетики независимо от физических особенностей объектов». Л. А. Растригин образно называет кибернетику «наукой о борьбе с хаосом», подчеркивая тем самым мысль о том, что управление противодействует дезорганизации и разрушению, вызываемым всевозможными случайными факторами.

Кибернетика и роботы. Один из центральных вопросов кибернетики — проблема *автоматизации процессов,* и в частности проблема *самоуправления (автоматизма) в сложных системах.* Исследования этих проблем привели к возникновению научно-технического направления, получившего название «робототехника». В современной кибернетической литературе обсуждаются возможности создания машин-автоматов, которые сами бы себя воспроизводи-

ли, были бы способны к самообучению. Исследуется проблема искусственного интеллекта. Ставятся вопросы: способна ли машина заниматься творческой деятельностью? Может ли машина стать умнее своего создателя? Может ли машина мыслить?

Роботы всевозможных типов перешагнули сегодня рамки науки и заполонили страницы научно-фантастических книг. Все чаще возникают дискуссии о перспективах робототехники и, в частности, о возможности создания искусственного человека. Неискушенные начинают полагать, будто кибернетика — это не что иное, как наука о роботах, удивительных автоматах, «думающих» машинах. Сущность кибернетики как науки об управлении оказывается при этом замаскированной эффектными техническими перспективами.

Кибернетика действительно исследует проблемы автоматизации, внося тем самым исключительный по важности вклад в научно-технический прогресс. Автоматизация многочисленных производственных процессов, создание лунохода, осуществление автоматической стыковки в космосе — все это входит в обширный перечень достижений кибернетики. Вместе с тем кибернетика исследует также возможности реализации машинного творчества, возможности создания искусственного интеллекта. Это делается не для того, чтобы в будущем изготовлять искусственных людей. Когда мы придумываем машины, «сочиняющие» музыку или рассказы, играющие в шахматы, «беседующие» на ту или иную тему, мы тем самым исследуем принципиально важную проблему моделирования творческих процессов, что позволяет глубже проникнуть в природу таких процессов. Можно сказать, что мы исследуем экстремальные возможности машин не с тем, чтобы в будущем заменить машинами живых людей, а с тем, чтобы лучше разобраться в ряде принципиально важных проблем, позволяющих понять процессы управления, происходящие в живом человеке. Читатель должен помнить об этом и не должен сводить кибернетику к «науке о роботах». Здесь самое время начать разговор о центральном понятии кибернетики — об *информации*. Сразу же подчеркнем, что кибернетика изучает процессы управления и самоуправления прежде всего с точки зрения информации. Она исследует вопросы возникновения, передачи, преобразования, хранения информации. В известном смысле кибернетика может рассматриваться как «наука об информации».

Информация

Начнем с отрывка из бессмертной поэмы Лукреция Кара «О природе вещей»:

Если б из ничего в самом деле являлися вещи,
Всяких пород существа безо всяких семян бы рождались.
Так, например, из морей возникали бы люди, из суши —
Рыб чешуйчатых род и пернатые, с неба срывался б
Крупный и мелкий скот, и породы бы диких животных
Разных, неведомо как, появлялись в полях и пустынях...

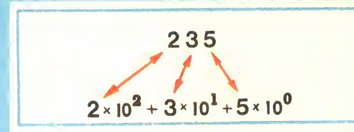
Любопытно, что здесь мы находим прозрачные намеки на сохранение не только вещества и энергии, но и чего-то еще, что не есть ни вещество, ни энергия. В море нет недостатка в энергии и веществе, однако люди не возникают из морской воды. Точно так же сама по себе суша не порождает рыб. И вообще, оказывается, нельзя, чтобы какие бы то ни было существа «безо всяких семян бы рождались». Если воспользоваться современной научной терминологией, то можно сказать, что здесь есть намек на *сохранение информации*. Богатая информация, содержащаяся в растениях, живых существах, не может возникнуть «из ничего». Она хранится в «семенах», передается по наследству.

Термин «информация» широко применяется в современной науке да и во всей человеческой практике. Фактически вся деятельность человека связана с *переработкой, получением, передачей, хранением информации*. Мы живем в мире, насыщенном разнообразной информацией, без нее само наше существование невозможно. Об этом хорошо сказал академик А. И. Берг: «Информация проникает во все по́ры жизни людей и обществ... Жизнь невозможна ни в вещественно-энергетическом, ни в информационном вакууме».

Бит — единица информации. Что такое информация? В каких единицах она измеряется? Начнем с простого примера. Поезд подходит к станции. Используя переключатель дистанционного управления железнодорожной стрелкой, диспетчер может направить поезд либо на путь *A*, либо на путь *B*. Если переключатель поставить в «верхнее положение», стрелка откроет путь *A*, а если в «нижнее» — путь *B*. Ставя переключатель в то или другое положение, диспетчер тем самым посылает управляющий сигнал, содержащий информацию в *1 бит*. Слово «бит» происходит от словосочетания *binary digit*, которое в переводе с английского означает «двоичная цифра».

Чтобы пояснить термин «двоичная цифра», напомним, как используются цифры для записи чисел. Обычно мы пользуемся *десятичной* системой счисления, т. е. системой с десятью цифрами (0, 1, 2, ..., 9). Возьмем какое-нибудь число. Пусть это будет число, которое выглядит в десятичной системе так: 235. Мы говорим «двести тридцать пять» и, как правило, не задумываемся над тем, что перед нами сумма двух сотен, трех десятков и пяти единиц: $2\times10^2+3\times10^1+5\times10^0$. То же самое число в *двоичной* системе счисления записывается при помощи всего лишь двух цифр (0 и 1); в этой системе данное число имеет следующий вид: 11101011. Эта запись расшифровывается так: $1\times2^7+1\times2^6+1\times2^5+0\times2^4+1\times2^3+0\times2^2+1\times2^1+1\times2^0$. Действительно, поскольку $2^7=128$, $2^6=64$, $2^5=32$, $2^3=8$, $2^1=2$, $2^0=1$, то последняя сумма есть не что иное, как $128+64+32+8+2+1=235$. Любое число можно записать как в десятичной, так и в двоичной системе. Если наши объяснения показались читателю не совсем понятными, советуем ему внимательно рассмотреть рисунок 3.3.

Вернемся к железнодорожной стрелке. У нас имеются, напоминаем, два исхода: переключатель в «верхнем положении» (открыт путь *A*),


235
2 × 10^2 + 3 × 10^1 + 5 × 10^0

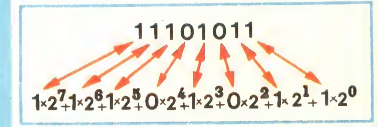

11101011
1×2^7+1×2^6+1×2^5+0×2^4+1×2^3+0×2^2+1×2^1+1×2^0

Рис. 3.3

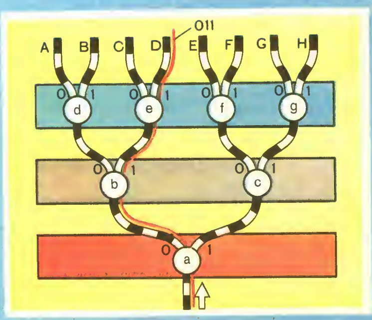

011
A
B
C
D
E
F
G
H
0
1
d
e
f
g
b
c
a

Рис. 3.4

переключатель в «нижнем положении» (открыт путь B). Закодируем первый исход цифрой 0, а второй — цифрой 1. Можно сказать, что управляющий сигнал кодируется в данном случае одной из двух двоичных цифр — либо нулем, либо единицей. Это и есть информация в одну двоичную цифру, или, короче, информация в 1 бит.

Рассмотрим более интересный пример. На рисунке 3.4 дана схема разветвления железнодорожных путей при подходе к станции. Железнодорожные стрелки обозначены буквами (a, b, c, d, e, f, g). Если стрелка получает управляющий сигнал 0, она открывает левый путь, а если получает сигнал 1, то открывает правый путь. Перед диспетчером три переключателя: первый подает сигнал (0 или 1) на стрелку a, второй подает сигнал одновременно на стрелки b и c, а третий — одновременно на стрелки d, e, f, g. На станции восемь путей: A, B, C, D, E, F, G, H. Чтобы направить поезд на путь A, надо поставить все три переключателя в положение 0, т. е. подать систему сигналов 000. Чтобы направить поезд на путь B, надо подать систему сигналов 001. Каждому пути отвечает свой набор сигналов:

A	*B*	*C*	*D*	*E*	*F*	*G*	*H*
000	001	010	011	100	101	110	111

Мы видим, что для выбора одного из восьми исходов нужен всякий раз набор из трех элементарных сигналов, каждый из которых несет информацию в 1 бит. Следовательно, всякий раз нужна информация в 3 бита.

Итак, чтобы выбрать один вариант из двух, нужен 1 бит информации; чтобы выбрать один вариант из восьми, нужны 3 бита информации. Для выбора одного из N вариантов нужна информация в I битов:

$$I=\log_2 N. \qquad (3.1)$$

Это есть *формула Хартли*. Она предложена в 1928 году американцем Хартли, интересовавшимся вопросами количественной оценки информации.

Игра «Бар-Кохба». В 135 году в древней Иудее вспыхнуло восстание против владычества римлян. Предводителем восставших был Бар-Кохба. Согласно легенде, Бар-Кохба послал в лагерь римлян лазутчика. Тот многое выведал, но был схвачен и брошен в темницу. Его пытали, вырвали ему язык. Лазутчику удалось бежать из темницы. Однако, не имея языка, он не мог рассказать о том, что подсмотрел во вражеском лагере. Но Бар-Кохба нашел выход из положения. Он стал задавать лазутчику такие вопросы, на которые достаточно было ответить лишь «да» или «нет» (достаточно кивнуть или покачать головой). Используя набор таких вопро-

сов, Бар-Кохба смог получить от безъязыкого лазутчика всю необходимую информацию.

В романе А. Дюма «Граф Монте-Кристо» описывается аналогичная ситуация. Один из героев романа, старик Нуартье, разбит параличом; он не может ни говорить, ни двинуть рукой. И тем не менее родственники общаются с ним, предлагая ему лишь такие вопросы, на которые требуется ответить «да» или «нет». В первом случае Нуартье закрывает глаза, а во втором несколько раз мигает.

Оказывается, что любую информацию можно представить в виде ответов «да» или «нет» на *соответствующим образом* сконструированные вопросы. Эта идея и лежит в основе *игры «Бар-Кохба»*, появившейся в начале нашего столетия сначала в Венгрии, а затем и других странах. Один игрок что-то загадывает. Он может, например, загадать некое желание и даже целую фразу. Другой должен выяснить, что задумал партнер. Для этого он задает партнеру различные вопросы, на которые тот честно отвечает. Существенно, чтобы задаваемые вопросы предполагали лишь ответы «да» или «нет». Количество информации, необходимое для отгадывания, можно измерить числом вопросов при наиболее рациональном способе ведения дознания. Каждый ответ можно закодировать одной из двоичных цифр. Можно, например, сопоставлять единицу с ответом «да», а нуль с ответом «нет». Тогда необходимая для отгадывания информация будет закодирована в виде некоторой комбинации единиц и нулей.

Сыграем в игру «Бар-Кохба» с диспетчером железнодорожной станции, для которой мы рассматривали схему путей на рисунке 3.4. Диспетчер загадал, на какой путь будет принят поезд, приближающийся к станции. Мы хотим выяснить, что загадал диспетчер. Игра может протекать так:

Вопрос: откроет ли стрелка *a* правый путь? Ответ: нет (кодируем этот ответ цифрой 0). Вопрос: откроет ли стрелка *b* правый путь? Ответ: да (кодируем: 1). Вопрос: откроет ли стрелка *e* правый путь? Ответ: да (кодируем: 1).

Задав три вопроса, мы выяснили, что диспетчер загадал путь *D*. Необходимая для отгадывания информация может быть закодирована цепочкой ответов «нет-да-да», или, иначе, набором двоичных цифр 011. Мы знаем, что информационная емкость диспетчерской «загадки» равна 3 битам. Каждый из трех ответов диспетчера содержал информацию в 1 бит.

Приведем еще один простой пример игры «Бар-Кохба». В классе 32 ученика. Учитель загадал одного из них. Как выяснить, кого именно? Возьмем классный журнал, где все фамилии учащихся расположены в алфавитном порядке и перенумерованы. Начнем задавать вопросы.

Вопрос: находится ли загаданный среди номеров с 17-го по 32-й? Ответ: да (кодируем: 1). Вопрос: находится ли он среди номеров с 25-го по 32-й? Ответ: нет (0). Вопрос: находится ли он среди номеров с 21-го по 24-й? Ответ: нет (0). Воп-

рос: находится ли он среди номеров 19 и 20? Ответ: да (1). Вопрос: стоит ли он под номером 20? Ответ: нет (0).

Итак, был загадан ученик, фамилия которого стоит в журнале под номером 19. Полученная информация кодируется цепочкой ответов «да-нет-нет-да-нет» или, иначе, набором двоичных цифр 10010. Из рисунка 3.5 видно, как постепенно уменьшается область поиска задуманной фамилии по мере получения ответа на очередной вопрос. Для решения задачи понадобилось задать пять вопросов. Согласно формуле Хартли, для выбора одного варианта из 32 нужна информация, равная $\log_2 32 = 5$ битам. Следовательно, каждый из полученных в данной игре ответов содержал информацию в 1 бит.

Может сложиться впечатление, что в игре «Бар-Кохба» каждый ответ всегда содержит информацию в 1 бит. Легко убедиться, что это не так. Допустим, что, установив присутствие загаданного лица среди номеров с 17-го по 32-й, мы затем задаем вопрос: находится ли это лицо среди номеров с 9-го по 16-й? Ясно, что на такой вопрос будет дан отрицательный ответ. Очевидность ответа означает, что он вообще не содержит информации. Можно, конечно, предложить ситуацию и без подобных заведомо «глупых» вопросов.

Вопрос: находится ли загаданный среди номеров с 1-го по 8-й? Ответ: нет. Вопрос: находится ли он среди номеров с 25-го по 32-й? Ответ: нет. Вопрос: находится ли он среди номеров с 9-го по 16-й? Ответ: нет. Вопрос: находится ли он среди номеров с 17-го по 24-й? Ответ: да. Вопрос: находится ли он среди номеров 17 и 18? Ответ: нет. Вопрос: находится ли он среди номеров 23 и 24? Ответ: нет. Вопрос: находится ли он среди номеров 19 и 20? Ответ: да. Вопрос: стоит ли он под номером 19? Ответ: да.

Выбрав такую стратегию, мы получаем нужную информацию, используя уже не пять, а восемь вопросов. Количество информации по-прежнему равно 5 битам. Значит, в данном случае один ответ содержал в среднем всего лишь 5/8 бита информации.

Итак, мы видим, что ответ «да-нет» не всегда содержит 1 бит информации. Забегая вперед, заметим, что 1 бит — это *максимальная* информация, которая может содержаться в таком ответе.

«Но позвольте,— может заметить читатель,— в таком случае и двоичная цифра не всегда несет информацию в один бит?» — «Совершенно верно»,— скажем мы. «Тогда как же быть с данным выше определением бита информации? И можно ли пользоваться формулой Хартли?» Все, что говорилось выше о бите информации (и о формуле Хартли), остается в силе, но с одной оговоркой: *варианты должны быть равновероятными.* Мы не хотели преждевременно говорить об этом. Теперь же настало время такого разговора.

Информация и вероятность. Формула Шеннона. Мы уже подчеркивали, что управление возможно лишь в мире, где необходимости диалектически противостоит случайность. Чтобы управлять, надо

иметь выбор. Ситуация, в которой мы хотим осуществить управление, должна нести в себе неопределенность. Неопределенность можно сопоставить с нехваткой информации. Осуществляя управление, мы вносим информацию и тем самым уменьшаем неопределенность.

Например, поезд можно принять на любой из восьми путей, налицо неопределенность. Посылая управляющий сигнал информационной емкостью в 3 бита, диспетчер ликвидирует эту неопределенность — поезд направляется на какой-то конкретный путь. Учитель мог загадать любого из 32 учеников — налицо неопределенность. Получив ответ на ряд вопросов общей информационной емкостью в 5 битов, мы ликвидируем эту неопределенность и выявляем избранника.

Но вернемся к исходной точке наших рассуждений — к вопросу о наличии выбора. До сих пор мы предполагали равновероятность выбора. Для нашего диспетчера был равновероятен выбор любого из восьми путей. Для учителя было все равно, кого именно загадать из 32 учеников. Часто приходится выбирать между вариантами, не являющимися равновероятными, и тогда необходимо принимать во внимание *вероятность выбора того или иного варианта.* Предположим, что задается вопрос, ответ на который имеет два исхода — «да» или «нет». Если оба исхода равновероятны, то ответ несет информацию ровно в 1 бит. Если же исходы «да» и «нет» имеют разную вероятность, то в ответе содержится информация меньше 1 бита, причем тем меньше, чем сильнее различаются вероятности исходов. В предельном случае, когда вероятность «да» (или «нет») обращается в единицу, ответ вообще не содержит информации.

Итак, будем полагать, что различные исходы (различные варианты) характеризуются разными вероятностями. Не желая загромождать книгу математическими выкладками, ограничимся обсуждением основных результатов. Пусть ξ — случайная дискретная величина, которая может принимать значения x_1, x_2, x_3, ..., x_N с вероятностями, равными соответственно p_1, p_2, p_3, ..., p_N. Перед нами N исходов (N разных значений случайной величины), которые реализуются с разными вероятностями. Мы производим наблюдение над величиной ξ и в результате выясняем, что она приняла такое-то значение. Сколько информации мы получаем в результате произведенного наблюдения?

Этот вопрос исследовал в середине 40-х годов нашего столетия американский ученый Клод Шеннон. Он пришел к выводу, что в рассматриваемой ситуации мы получаем количество информации, равное (в битах)

$$I(\xi)=\sum_{i=1}^{N} p_i \log_2 \frac{1}{p_i}. \quad (3.2)$$

Это соотношение является одним из основных в теории информации. Его называют *формулой Шеннона.*

ПОСЛЕ ОТВЕТА —

17 32

на первый вопрос

17 24

на второй вопрос

17 20

на третий вопрос

19 20

на четвертый вопрос

19

на пятый вопрос

Рис. 3.5

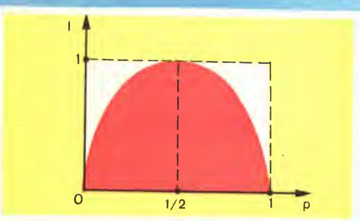


Рис. 3.6

Предположим, что исходы равновероятны — значения x_i случайной величины реализуются с одинаковой вероятностью p. Эта вероятность равна, очевидно, $1/N$. В данном случае получаем из (3.2):

$$I=\frac{1}{N}\sum_{i=1}^{N}\log_2 N=\frac{1}{N}N\log_2 N=\log_2 N,$$

т. е. приходим к формуле Хартли (3.1). Мы видим, таким образом, что формула Хартли получается из формулы Шеннона как *частный* случай — когда все исходы равновероятны.

Используя формулу Шеннона, выясним, сколько информации может содержаться в ответе на вопрос с двумя возможными исходами («да» или «нет»). Пусть p — вероятность того, что на поставленный вопрос будет дан ответ «да». Тогда вероятность ответа «нет» есть $1-p$. Согласно (3.2), информация, получаемая из ответа на рассматриваемый вопрос, равна

$$I=p\log_2\frac{1}{p}+(1-p)\log_2\frac{1}{1-p}. \qquad (3.3)$$

На рисунке 3.6 дан график зависимости I от p, определяемой соотношением (3.3). Видно, что максимум информации (1 бит) получается тогда, когда $p=1/2$, т. е. когда исходы «да» и «нет» равновероятны. Настало время уточнить понятие «1 бит информации». *Это есть информация, содержащаяся в кодовом знаке, принимающем лишь два значения, при условии, что оба эти значения равновероятны.*

Отсюда следует, что наилучшая стратегия в игре «Бар-Кохба» та, при которой вопросы предполагают ответы с равновероятными (или почти равновероятными) «да» и «нет». Напомним вопрос: находится ли загаданное лицо среди номеров с 17-го по 32-й? Ответы «да» и «нет» здесь равновероятны, так как всего имеется 32 ученика и номера с 17-го по 32-й охватывают ровно половину учеников. Поэтому ответ на данный вопрос дает ровно 1 бит информации. А вот иной вопрос: находится ли загаданное лицо среди номеров с 1-го по 8-й? Указанный интервал номеров охватывает четверть всех номеров, поэтому вероятность «да» равна 1/4, а «нет» 3/4. Ответ на поставленный вопрос содержит информацию меньше бита. Согласно (3.3), где надо принять $p=1/4$, эта информация равна 0,8 бита.

Еще раз подчеркнем мысль о том, что процессы управления должны рассматриваться в диалектическом единстве со случайными процессами дезорганизации. Уже отсюда можно сделать вывод о глубокой связи между теорией информации и теорией вероятностей. Формула Шеннона (3.2) наглядно демонстрирует эту связь. Именно *вероятностный* подход дает научное, объективное понятие информации, свободное от субъективных представлений, подменяющих количество информации ее ценностью и важностью.

Передача информации по каналу связи с помехами. При передаче информации неизбежны ее *потери*. Они объясняются действием

различных случайных факторов. Их обычно называют *помехами.* На рисунке 3.7 схематично изображен некий *канал связи,* по которому передается информация — от входа A к выходу B. В процессе передачи по каналу на информацию воздействуют помехи P. Предположим, что на вход A поступает случайная дискретная величина ξ, значения которой $x_1, x_2, ..., x_N$ реализуются с вероятностями $p_1, p_2, ..., p_N$. Предположим далее, что на выходе B мы принимаем величину η, значения которой $y_1, y_2, ..., y_M$ реализуются с вероятностями $q_1, q_2, ..., q_M$. Обозначим через $P_i(j)$ вероятность наблюдения на выходе значения $\eta = y_j$ при условии, что на вход подано значение $\xi = x_i$. Вероятность $P_i(j)$ определяется помехами в канале связи. В теории информации доказывается, что количество информации о случайной величине ξ, получаемое при наблюдении случайной величины η, описывается формулой:

$$I_\eta(\xi) = \sum_{i=1}^{N} \sum_{j=1}^{M} P_i(j) p_i \log_2 \frac{P_i(j)}{q_j} . \qquad (3.4)$$

Здесь информация I выражается через вероятности двух видов — вероятности p_i и q_j и вероятность $P_i(j)$. Если первые две вероятности отражают вероятностную природу информации, которая соответственно подается на вход канала связи и принимается на его выходе, то вероятность $P_i(j)$ отражает случайный характер помех в канале связи.

Предположим, что помехи отсутствуют. Тогда будет наблюдаться однозначное соответствие между значениями случайной величины, подаваемой на вход канала, и значениями случайной величины, принимаемой на выходе. Это значит, что

$$N = M;\ p_i = q_i;\ P_i(j) = \delta_{ij}, \text{ где } \delta_{ij} = 1 \text{ при } i = j \text{ и } \delta_{ij} = 0 \text{ при } i \neq j. \qquad (3.5)$$

Подставив (3.5) в (3.4) и учитывая, что $\lim\limits_{z \to 0} z \log_2 z = 0$, получим формулу Шеннона. Этого следовало ожидать, поскольку в отсутствие помех нет и потери информации в процессе передачи.

Борьба с помехами в канале связи. Каналы связи весьма разнообразны. Информация может передаваться при помощи звуковых волн, распространяющихся по воздуху или в иной среде; электрических сигналов, бегущих по проводам; электромагнитных волн, распространяющихся в среде или в вакууме, и т. д. Каждый канал связи имеет специфические помехи. Существуют общие способы борьбы с помехами, пригодные для различных каналов связи. Прежде всего желательно максимально снижать уровень помех и максимально повышать уровень полезного сигнала. Как говорят, надо увеличивать отношение сигнал/шум. Увеличение этого отношения может достигаться также за счет соответствующего кодирования передаваемой информации, т. е. представления ее в виде таких «символов» (например, импульсов определенной формы), которые четко выделялись бы на фоне помех. Такое кодирование повышает *помехоустойчивость* передаваемой информации.



Рис. 3.7

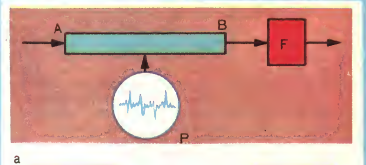


Рис. 3.8

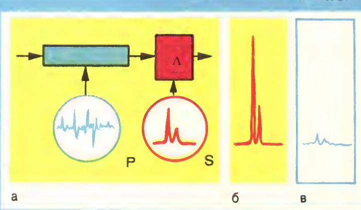


Рис. 3.9

Особое место в борьбе с помехами занимает *фильтрация* информации, принимаемой на выходе канала связи. Различают *осредняющую* и *корреляционную* фильтрацию. Предположим, что характерная частота шумов в канале связи существенно выше частоты, характеризующей изменения во времени полезного сигнала. В этом случае можно поставить на выходе канала связи *осредняющий фильтр*, который попросту «срежет» высокочастотные колебания, наложившиеся на полезный сигнал в процессе его передачи по каналу. Сказанное поясняет рисунок 3.8, где *а* — схема канала связи с фильтром (*A* — вход канала, *B* — выход канала, *P* — помехи, *F* — осредняющий фильтр), *б* — сигнал, подаваемый на вход канала, *в* — сигнал на выходе до фильтрации, *г* — сигнал после фильтрации.

Предположим, что надо выяснить, присутствует ли в принимаемой на выходе канала информации некий сигнал, форма которого известна. Если этот сигнал существенно отличается (например, по частоте) от сигналов, связанных с помехами, то такое выяснение, или, как обычно говорят, *распознавание*, особого труда не составит. Хуже, когда распознаваемый сигнал маскируется шумами. В этих случаях применяют корреляционную фильтрацию: на выходе канала связи помещают специальное устройство, выполняющее *перемножение* выходного сигнала и искомого сигнала. Если искомый сигнал действительно присутствует в выходном сигнале, перемножение приведет к появлению *интенсивного сигнала корреляции*; в противном случае сигнал корреляции не появится. Сказанное поясняет рисунок 3.9, где *а* — схема канала связи (Λ — устройство для перемножения сигналов, *P* — помехи, *S* — опознаваемый сигнал), *б* — сигнал после перемножения в случае, когда опознаваемый сигнал *S* присутствует на выходе канала (сигнал корреляции), *в* — сигнал после перемножения в случае, когда сигнал *S* отсутствует на выходе канала. Корреляционная фильтрация применяется, в частности, в *радиолокации* — для распознавания посланного сигнала в излучении, принимаемом радиолокатором.

Отбор информации из шума

Проблема возникновения информации и неудачные подходы к ее решению. Управляющий сигнал несет в себе определенную информацию. Формирование сигнала происходит по определенному алгоритму, содержащему соответствующую информацию. В свою очередь, указанный алгоритм составлен на основе информации, заложенной в других алгоритмах управления, тех, которые были использованы при составлении данного алгоритма. Возникает мысль о своеобразной эстафете, связанной с передачей информации от одних алгоритмов к другим. Эту мысль можно пояснить на очень простом примере. Учитель обучает вас; в свою очередь, ваш учитель должен был у кого-то учиться; этот кто-то имел своих учителей. И так далее.

Подобные рассуждения почти неизбежно порождают вопросы: от-

куда взялась «изначальная информация»? Откуда появились первые алгоритмы? Неумение (или нежелание) научно исследовать принципиально важную проблему возникновения информации приводит к серьезным заблуждениям.

Одно из таких заблуждений заключается в том, что акт творения «изначальной информации» приписывают богу. Это заблуждение представляет собой идеологическую основу религии. Порывая с материализмом, она связывает создание информации с таинственной потусторонней волей «творца», который «придумал» и «сотворил» наш мир.

Другое заблуждение связано с гипотезой, согласно которой «изначальная информация» была принесена на Землю космическими пришельцами, якобы посетившими нас в отдаленные времена. Эта гипотеза не порывает с материализмом, однако она также не в состоянии решить проблему. Ведь остается открытым вопрос о том, откуда взяли необходимую информацию сами пришельцы.

Современная наука решила проблему возникновения информации. Согласно современным научным представлениям, никакой «изначальной информации» вообще не существовало — процесс возникновения (генерации) информации является непрерывным, всегда совершающимся процессом.

На сцену опять выходит случайность. Представление, будто информация передается просто как своеобразная эстафетная палочка,— упрощенное. Выше мы уже говорили о том, что всякая передача информации сопровождается потерями, которые обусловлены действием случайных факторов. Однако случайности не только «крадут» информацию, они являются и *генератором информации.*

На первый взгляд, высказанная мысль кажется неправдоподобной. Все мы свидетели непрекращающегося процесса создания все новой и новой информации в результате творческой деятельности людей. Новые машины сходят с конвейеров, новые космические корабли выводятся на орбиты, новые книги выходят из печати, новые лекарства появляются в аптеках — все это свидетельствует о бурном процессе генерации информации, в котором все мы принимаем то или иное участие. И как-то странно на фоне всего этого говорить о принципиальной роли случая как генератора информации.

Однако задумайтесь, например, каким образом совершается *процесс мышления,* как рождается решение какой-либо задачи, как появляется новая идея, как возникает мелодия или художественный образ. Попытайтесь понять хотя бы механизм *ассоциативного восприятия,* помогающий нам узнавать предметы, различать их. Попробуйте — и вы вступите в область сложнейших связей, вероятностных отношений, случайных догадок, внезапных «озарений». Не существует детерминированных алгоритмов для того, чтобы делать открытия и решать проблемы. Все, что мы знаем сегодня о процессах, происходящих в нашем мозгу, указывает на *принципиальную роль* случайных факторов в этих про-

цессах. Позднее мы продемонстрируем это на примере *персептрона* — кибернетического устройства, способного к распознаванию образов.

Случай и отбор. Каким же образом случайность может генерировать информацию? Каким образом из беспорядка может возникать порядок? Оказывается, возникновение информации из шума совсем нетрудно наблюдать. Читатель может убедиться в этом. Воспользуемся известной игрой под названием «Эрудит», а точнее, применяемыми в этой игре квадратиками с изображениями различных букв. Отберем по одной букве весь алфавит и ссыпем отобранные квадратики в мешочек. Перемешаем их и будем наугад вынимать. Каждую случайным образом вынутую букву будем записывать, а квадратик будем возвращать обратно. Всякий раз надо тщательно перемешивать квадратики. Используя этот простой генератор случайных букв, мы можем выписать сколь угодно длинную хаотическую последовательность различных букв. И вот если вы внимательно посмотрите на нее, то обнаружите отдельные трехбуквенные слова. Могут встретиться и слова из четырех и более букв. Налицо акт рождения информации из шума!

Прибегнув к помощи сына, автор сам проделал такой опыт. В последовательности из 300 случайных букв он обнаружил 9 трехбуквенных слов и 2 четырехбуквенных. Чем больше букв в слове, тем меньше вероятность возникновения слова в «буквенном шуме». Еще менее вероятно рождение целой фразы или тем более строки из известного произведения. И тем не менее все эти вероятности не равны нулю, так что существует принципиальная возможность того, что из шума случайно возникнет фактически любая информация.

Итак, мы можем сказать (хотя это звучит как каламбур), что *случайности порождают информацию случайно*. При этом, чем больше информация, тем меньше вероятность ее случайного возникновения. Сам факт случайного возникновения той или иной информации еще не решает проблемы. Необходимо выделить эту неожиданно возникшую информацию из огромного потока бессмысленных «сигналов». Иначе говоря, необходимо произвести *отбор информации из шума*. В примере с выниманием букв-квадратиков отбор информации из шума производит человек, который сначала фиксирует все выпавшие буквы на бумаге, а затем просматривает получившуюся запись.

Усилитель отбора. А нельзя ли сознательно использовать случайность для генерации информации? Можно, если только позаботиться об *усилении отбора*.

Простой опыт, демонстрирующий усиление отбора, читатель может проделать, используя генератор случайных букв, описанный выше. Чтобы усилить отбор, надо положить в мешочек не один, а несколько экземпляров каждой буквы — с учетом частоты появления той или иной буквы в словах. Автор проделал такой опыт, прикинув частоту появления тех или иных букв «на глазок». Он вообще изъял буквы Й, Ъ, Э, Ю, Ц, зато буквы А и О использовал в

пяти экземплярах, буквы И и Е — в четырех, буквы Л, М, Р, С, Т — в трех, буквы Г, Д, К, Н, П, У — в двух, остальные буквы — в одном экземпляре. Нельзя ручаться, что такой выбор разных букв был оптимальным, однако эффект усиления отбора оказался налицо — в последовательности из 300 случайных букв было обнаружено 21 трехбуквенное слово, 4 четырехбуквенных и 1 пятибуквенное.

Чтобы еще более усилить отбор, следовало бы использовать не отдельные *буквы*, а отдельные *слова*. Любопытно, что соответствующее устройство было предложено еще в первой половине XVIII века. Его описал английский писатель Джонатан Свифт в своей знаменитой книге о путешествиях Гулливера. Когда Гулливер посетил Академию в Лагадо (столица фантастического королевства), он познакомился там с устройством из множества нанизанных на спицы кубиков, на сторонах которых были написаны «все слова их языка в различных наклонениях, временах и падежах, но без всякого порядка». По команде местного профессора ученики несколько раз повертывали спицы с кубиками, что приводило к изменениям сочетаний слов. «Тогда профессор,— пишет Свифт,— приказал тридцати шести ученикам медленно читать образовавшиеся строки в том порядке, в каком они разместились в раме. Если случалось, что три или четыре слова составляли часть фразы, то ее диктовали остальным четырем ученикам, исполнявшим роль писцов. Это упражнение было повторено три или четыре раза, и машина была так устроена, что после каждого оборота спиц слова принимали новое расположение по мере того, как кубики переворачивались с одной стороны на другую». Правда, Свифт пишет обо всем этом в сатирическом духе, он как бы высмеивает подобные изобретения. Однако почему бы нам не считать, что тут под маской сатирика скрывался талантливый фантаст, который прибег к сатире из опасения быть непонятым и осмеянным современниками?

И вот то, что казалось смешным и нелепым в XVIII веке, превратилось в объект научного исследования в середине XX века. В начале 50-х годов нашего столетия английский ученый У. Росс Эшби предложил создать кибернетическое устройство, представляющее собой *усилитель отбора*. Эшби назвал его *усилителем мыслительных способностей*. Схема этого усилителя дана на рисунке 3.10. Генератор шума *1* поставляет «сырье» в первую ступень усилителя. Преобразователь шума *2* создает различные случайные варианты объектов отбора. В блоке *3* происходит отбор в соответствии с заложенными в устройство критериями отбора. Если результат отбора в том или ином конкретном случае удовлетворяет критерию, срабатывает блок управления *4*, открывая клапан *5* и пропуская отобранную информацию в преобразователь следующей ступени усилителя. Можно представить себе, что в первой ступени усилителя, куда поступают случайные буквы, происходит отбор отдельных случайно возникших слов или отдельных характерных слогов; во второй ступени усилителя проис-

4 Зак. 661 Тарасов

ходит отбор сочетаний слов; в третьей ступени — отбор целых фраз; в четвертой — мыслей и т. д.

Самоорганизация на основе случайного поиска. Гомеостат. Пусть некая система находится в каком-то состоянии, которое позволяет ей выполнять определенные функции. Условно назовем такое состояние *нормальным*. Оно соответствует внешним условиям, в которых работает система. Предположим, что условия вдруг изменились, в результате чего система вышла из нормального состояния. Новым условиям соответствует и новое нормальное состояние. Желательно перевести систему в это новое состояние. Как это сделать? Нужна информация, во-первых, о новом состоянии и, во-вторых, о том, как может быть осуществлен перевод системы в это состояние. Поскольку изменение внешних условий имеет, как правило, случайный характер, поэтому мы не знаем, ни каким будет новое нормальное состояние, ни как организовать переход в него. В такой ситуации на помощь приходит *случайный поиск*. Это означает, что надо случайным образом изменять параметры системы до тех пор, пока она случайно не окажется в новом нормальном состоянии, о чем можно немедленно узнать, непрерывно контролируя поведение системы.

Можно сказать, что в процессе случайного поиска возникает как раз та информация, которая необходима для перевода системы в новое состояние. Это есть не что иное, как уже знакомый нам *отбор информации из шума*. Критерием отбора является здесь изменение поведения системы: попав в новое нормальное состояние, система «успокаивается», начинает нормально функционировать.

В 1948 году Эшби сконструировал прибор, который обладал свойством самоорганизации на основе случайного поиска. Он назвал его *гомеостатом*. Принципиальная схема гомеостата Эшби показана на рисунке 3.11. Основой прибора служит система *1*, которая может находиться либо в устойчивом, либо в неустойчивом состоянии. Не входя в технические детали, заметим лишь, что система *1* состоит из четырех электромагнитов, сердечники которых могут поворачиваться и при этом передвигать ползунки реостатов, управляющих питанием электромагнитов. Таким образом, углы поворота четырех магнитов оказываются взаимосвязанными. Эти углы и представляют собой параметры данной динамической системы. В устойчивом состоянии все магниты неподвижны. Пусть внешнее возмущение выводит гомеостат из устойчивого состояния. Немедленно блок управления *2* включает генератор случайных изменений параметров *3*. Начинается случайный поиск. Как только система *1* случайно попадает в устойчивое состояние, блок проверки устойчивости *4* подает сигнал и блок управления *3* выключает генератор случайных изменений параметров.

Гомеостат часто сравнивают со спящей кошкой. Если кошку потревожить, она проснется, выберет новое удобное положение и снова заснет. Так же ведет себя и гомеостат: «проснувшись», он осуществляет случайный поиск новых значений своих параметров и, найдя их, как бы снова «засыпает».

На пути к стохастической модели мозга

Проблема распознавания образов. Обычно мы не задумываемся над способностью мозга *распознавать* образы. А между тем способность эта удивительна. На рисунке 3.12 показаны несколько «фигур», различающихся размерами, толщиной линий, формой. Несмотря на это, мы сразу же узнаем во всех этих «фигурах» один и тот же образ — букву «А». Еще более удивительно, когда в толпе разнообразно одетых людей с плохо различимыми из-за расстояния лицами нам удается, как правило, безошибочно распознать, кто есть женщина, а кто мужчина.

Способность распознавать образы связывают с *ассоциативным восприятием* — когда воспринимаются некие общие, характерные признаки, а конкретные, частные признаки как бы отступают на второй план. Возможно ли ассоциативное восприятие у машины? Можно ли смоделировать происходящие в мозгу процессы, связанные с распознаванием образов? Положительный ответ на этот вопрос был получен в 1960 году, когда американский ученый Ф. Розенблатт сконструировал устройство, названное им *персептроном* (от латинского *perceptio*, что означает «понимание»).

Что такое персептрон? Персептрон можно рассматривать как сильно упрощенную модель системы «глаз-мозг». Роль глаза, а точнее сетчатки глаза, выполняет экран (табло), составленный из большого числа *светочувствительных элементов*, или *рецепторов*. На экран, как на сетчатку глаза, проецируется изображение. Каждый рецептор преобразует падающий на него свет в электрические сигналы, которые поступают внутрь персептрона к элементам, выполняющим анализирующие и решающие функции.

Прежде чем рассматривать устройство персептрона, отметим два принципиальных обстоятельства. Во-первых, связи между рецепторами и внутренними элементами персептрона, предназначенными для обработки информации, зафиксированной рецепторами, *не должны быть детерминированными*. Если бы они были детерминированными, то сигналы от изображений, показанных на рисунках 3.13, *а* и 3.13, *б*, «воспринимались» бы персептроном как явно разные образы (на этих рисунках совпадают только пять возбужденных рецепторов, они выделены красным цветом), а изображения на рисунках 3.13, *а* и 3.13, *в* «воспринимались» бы, напротив, как один образ — ведь здесь совпадают 28 возбужденных рецепторов. В действительности же персептрон должен «воспринимать» изображения на рисунках 3.13, *а* и 3.13, *б* как единый образ, а на рисунках 3.13, *а* и 3.13, *в* как разные образы. Итак, надо признать, что внутренние связи в персептроне должны быть реализованы *случайным образом*. Это должны быть *вероятностные* связи.

Во-вторых, случайный характер связей предполагает своеобразную *настройку* персептрона на распознаваемые образы. Персептрону надо предъявить поочередно (и по нескольку раз) различные изображения распознаваемых образов и провести своеоб-

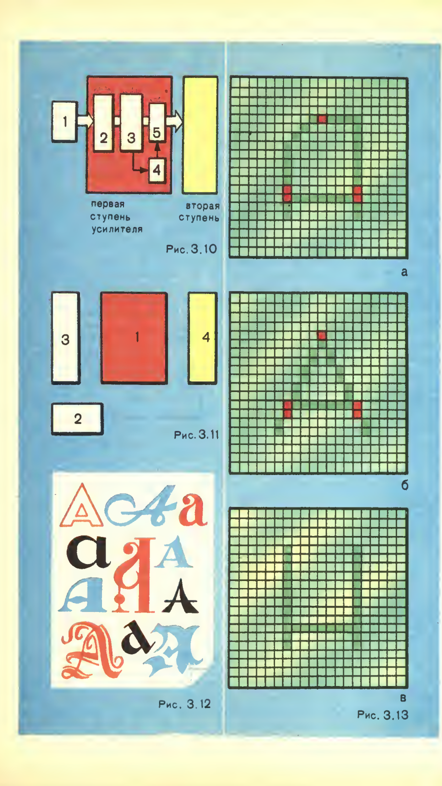


Рис. 3.10

Рис. 3.11

Рис. 3.12

Рис. 3.13

разное *обучение*, в процессе которого и произойдет необходимая настройка параметров персептрона. На каждом этапе (при каждом предъявлении изображения) необходимо сопоставление с ранее достигнутыми успехами, следовательно, персептрон должен обладать *памятью*.

Учитывая оба отмеченных обстоятельства, можем определить персептроны как устройства определенного типа, характеризующиеся наличием памяти и случайной структурой связей между элементами. Персептрон может рассматриваться как упрощенная модель мозга, перспективность этой модели связана с тем, что она является *вероятностной*, или, иначе говоря, *стохастической*. По мнению ученых, именно стохастические модели способны наиболее удачно представить сущность процессов, происходящих в мозгу.

В настоящее время разработаны персептроны различного типа. Ниже мы рассмотрим подробнее устройство персептрона наиболее простого типа, который позволяет различать всего два образа.

Как устроен простейший персептрон? Схема персептрона дана на рисунке 3.14. Здесь: S_i— светочувствительные элементы (рецепторы), I_k— инверторы, изменяющие знак электрического потенциала, A_j — ассоциативные элементы, или A-элементы, λ_j — усилители с изменяющимся коэффициентом усиления, Σ — сумматор, R — реагирующий элемент. Пусть полное число рецепторов S_i равно N ($i=1, 2, 3, ..., N$). В первых моделях оно составляло $20\times20=400$. Число инверторов неопределенно: оно различно в разных экземплярах одного и того же прибора. Полное число ассоциативных элементов A_j, а также усилителей λ_j равно M ($j=1, 2, ..., M$). От рецепторов к A-элементам идут провода. Соединения рецепторов с A-элементами могут быть двух типов — непосредственные и через инверторы. Существенно, что выбор рецепторов, подключаемых к данному A-элементу, так же как и выбор знака подключения, осуществляется случайно: при монтаже каждой конкретной схемы провода, соединяющие рецепторы с A-элементами, припаиваются *случайным образом*, например в соответствии с указаниями, поступающими от какого-нибудь генератора случайных чисел.

Предположим, что на световое табло персептрона спроецировано некоторое изображение. В зависимости от степени освещенности различных участков табло одни рецепторы окажутся возбужденными, другие останутся невозбужденными. В первом случае на выходе рецептора будем иметь электрический сигнал 1, а во втором — 0. При наличии инвертора сигнал 1 преобразуется в -1. По системе случайных связей сигналы от рецепторов поступают к A-элементам. Каждый A-элемент производит алгебраическое сложение сигналов, поступивших на его вход. Если при этом сумма оказывается выше некоторого определенного значения (выше некоторого порога), то на выходе A-элемента появляется сигнал $+1$, в противном случае имеем сигнал 0. Будем обозначать эти сигналы через y_j. Каждый y_j равен либо $+1$, либо 0. Сигнал с выхода

элемента A_j поступает на вход усилителя λ_j, усилитель преобразует сигнал y_j в сигнал $\varkappa_j y_j$. Коэффициент усиления $\varkappa_j$ можно варьировать не только по модулю, но также и по знаку. Суммарный сигнал от всех усилителей реализуется в сумматоре Σ, это есть сигнал

$$\sum_{j=1}^{M} \varkappa_j y_j.$$

Он поступает затем на вход R-элемента, который проверяет его знак. Если окажется, что $\sum_j \varkappa_j y_j \geqslant 0$, то R-элемент формирует сигнал $+1$, а если $\sum_j \varkappa_j y_j < 0$, то на выходе R-элемента возникнет сигнал 0.

Рассматриваемый персептрон предназначен для распознавания всего двух образов. На один из образов (независимо от конкретного изображения этого образа) персептрон должен реагировать выходным сигналом $+1$, а на другой — сигналом 0. Чтобы персептрон приобрел такую способность, его надо *обучить*.

Обучение персептрона. Обозначим распознаваемые образы как B и C. Пусть образу B соответствует выходной сигнал $+1$, а образу C — сигнал 0. Пусть $\varkappa_1$, $\varkappa_2$, ..., $\varkappa_j$, ..., $\varkappa_M$ — последовательность коэффициентов усиления персептрона перед началом его обучения. Будем обозначать эту последовательность так: $\{\varkappa\}$. Начиная обучение, предъявляем персептрону первое изображение образа B. В результате возбудится некоторая совокупность A-элементов, т. е. появится некоторая последовательность сигналов y_1, y_2, ..., y_j, ..., y_M, или, короче, $\{y\}$. Допустим, что при этом сумма $\sum_j \varkappa_j y_j$ оказывается неотрицательной, так что выходной сигнал персептрона есть $+1$. Если это так, то пока все в порядке и можно предъявить персептрону второе изображение образа B. Второму изображению будет соответствовать новая совокупность возбужденных A-элементов, т. е. новая последовательность сигналов $\{y'\}$. Набор же коэффициентов усиления $\{\varkappa\}$ остается пока прежним. Сумма $\sum_j \varkappa_j y'_j$ может оказаться отрицательной, и тогда на выходе персептрона будет сформирован сигнал 0. Это нехорошо, поэтому персептрон получает своеобразное «наказание»: коэффициенты усиления возбужденных A-элементов увеличивают, скажем, на единицу, с тем чтобы новый набор коэффициентов усиления $\{\varkappa'\}$ обеспечил неотрицательность суммы $\sum_j \varkappa'_j y'_j$. Теперь персептрон правильно реагирует на второе изображение образа B. Однако как обстоят дела с первым изображением? Ведь набор коэффициентов усиления изменен, так что знак суммы $\sum_j \varkappa'_j y_j$ может быть любым. Предъявляем персептрону снова первое изображение образа B и по выходному сигналу выявляем знак суммы $\sum_j \varkappa'_j y_j$.

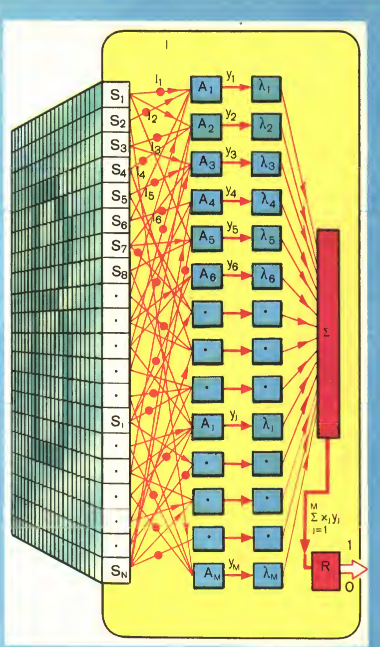

I
S1
S2
S3
S4
S5
S6
S7
S8
Si
SN
A1
A2
A3
A4
A5
A6
Aj
AM
y1
y2
y3
y4
y5
y6
yj
yM
λ1
λ2
λ3
λ4
λ5
λ6
λj
λM
Σ
R
1
0

Рис. 3.14

Если эта сумма оказывается неотрицательной, то мы удовлетворены — при наборе коэффициентов усиления $\{\varkappa'\}$ персептрон правильно реагирует как на первое, так и на второе изображение образа B. Можно теперь предъявлять персептрону третье изображение образа B. Если же указанная сумма оказалась отрицательной, то надо опять увеличить на единицу коэффициенты усиления возбужденных A-элементов (перейти от набора $\{\varkappa'\}$ к новому набору $\{\varkappa''\}$) и так далее.

Постепенно, изменяя шаг за шагом совокупность коэффициентов усиления, можно в конце концов найти такой набор этих коэффициентов, при котором персептрон будет выдавать сигнал $+1$ на любое предъявленное ему изображение образа B. Однако задача еще не решена. Вполне возможно, что после многократных увеличений различных коэффициентов усиления персептрон будет выдавать сигнал $+1$ не только на всевозможные изображения образа B, но также и на изображения образа C. А надо, чтобы, выдавая сигнал $+1$ на все изображения образа B, персептрон в то же время выдавал сигнал 0 на все изображения образа C. Значит, при обучении персептрона надо чередовать изображения обоих образов. Предъявляя изображение образа C, надо (при необходимости) уже не увеличивать, а уменьшать на единицу коэффициенты усиления возбужденных в данный момент A-элементов, с тем чтобы сумма $\Sigma \varkappa y$ оказалась отрицательной.

В итоге можно найти такой набор коэффициентов усиления $\{\varkappa^0\}$, при котором персептрон всегда будет распознавать образы B и C. Пусть $\{y(n)\}$ — набор возбужденных A-элементов, отвечающий n-му изображению образа B, а $\{Y(m)\}$ — набор, отвечающий m-му изображению образа C. Коэффициенты усиления $\{\varkappa^0\}$ должны быть такими, чтобы $\sum_{j=1}^{M} \varkappa_j^0 y_j(n) \geqslant 0$ при всех n и $\sum_{j=1}^{M} \varkappa_j^0 Y_j(m) < 0$ при всех m. Обучение окончено, как только такие коэффициенты усиления найдены.

Заключение. Заканчивая главу, подчеркнем главное, о чем здесь шла речь,— глубокую, внутреннюю связь между *теорией информации* и *теорией вероятностей*. Само понятие информации строится на основе вероятности. И это естественно, поскольку процессы управления и случайные процессы всегда выступают в *диалектическом единстве*. Случайность не только «крадет» информацию, но и генерирует ее, так как наиболее сложные информационные устройства принципиально основаны на случайной структуре внутренних связей.

Мы убеждаемся, что словосочетание «мир, построенный на вероятности», вынесенное в заглавие данной книги, приобретает все более глубокое содержание. Человек живет и активно действует в мире, насыщенном информацией. Информация же по самой своей природе проявляется через вероятность и, более того, создается в вероятностных процессах. Тем самым *мир, насыщенный информацией*, естественным образом оборачивается *миром, построенным на вероятности*.

Часть II

Фундаментальность вероятностных законов



ГЛАВА 4 Вероятность в классической физике

Первое применение теории вероятностей в физике, имеющее фундаментальное значение для нашего понимания законов природы, мы находим в основанной Больцманом и Гиббсом общей статистической теории теплоты...
Самое красивое и важное достояние этой теории — это понимание термодинамической «необратимости» как картины перехода к более вероятным состояниям.

В. Паули

Термодинамика и ее загадки

Все тела состоят из молекул, находящихся в хаотическом тепловом движении. Это принципиально важное обстоятельство можно не принимать во внимание при рассмотрении задач, составляющих предмет *термодинамики* — раздела физики, где с наиболее общих позиций (без использования молекулярных представлений) исследуются процессы обмена энергией между изучаемым макроскопическим объектом и окружающей его средой. Термодинамическое рассмотрение основано на описании состояний объекта при помощи некоторых специальных величин, называемых *термодинамическими параметрами,* и на использовании нескольких основных законов, называемых *началами термодинамики.* Хорошо известными читателю термодинамическими параметрами являются *температура* и *давление.*

Термодинамическое равновесие. Проделаем простой опыт. Внесем в комнату сосуд с нагретой водой и опустим в воду термометр. Следя за показаниями термометра, мы сможем убедиться, что температура воды в сосуде постепенно падает и в конечном счете становится равной температуре воздуха в комнате, после чего температура остается неизменной. Это означает, что вода в сосуде пришла в *термодинамическое (тепловое) равновесие* с окружающей ее средой. Если система находится в термодинамическом равновесии, ее термодинамические параметры (температура и давление) остаются неизменными сколь угодно долго. Другая характерная черта термодинамического равновесия — постоянство значений температуры во всех точках системы.

Если система не обменивается энергией с окружающими телами, то ее называют *замкнутой.* Говоря о термодинамическом равновесии замкнутой системы, подразумевают равновесие между частями системы, каждая из которых может рассматриваться как макроскопическое тело.

Предположим, что мы неравномерно нагрели некий объект, а затем поместили его в оболочку, не проводящую теплоты. Можно ска-

зать, что сначала мы нарушили в объекте термодинамическое равновесие, а затем предоставили объект самому себе. Температура более нагретых частей объекта будет уменьшаться, а холодных возрастать, и в конечном счете во всех частях объекта температура станет одинаковой — все части объекта придут в термодинамическое равновесие друг с другом. *Предоставленная самой себе макросистема всегда приходит в состояние термодинамического равновесия и остается в нем неопределенно долго* — до тех пор, пока какое-либо внешнее воздействие не выведет ее из этого состояния. Если затем указанное воздействие прекратится, то система снова придет в состояние термодинамического равновесия.

И вот вам первая загадка термодинамики. Почему выведенная из теплового равновесия система сама по себе возвращается в равновесное состояние, а система, находящаяся в состоянии теплового равновесия, сама по себе из него не выходит? Почему для поддержания теплового равновесия не нужно расходовать энергию, тогда как для удержания системы в термодинамически неравновесном состоянии необходимы затраты энергии? Вопрос этот, кстати говоря, далеко не праздный. На улице мороз −20°С, а в комнате тепло: 25°С. Стены современных домов довольно неплохо проводят теплоту. И, следовательно, налицо неравновесная система «комната+улица». Чтобы поддерживать это термодинамически неравновесное состояние, приходится непрерывно затрачивать энергию на отопление.

Первое начало термодинамики. Исследуемый объект (система) может обмениваться энергией с окружающей его средой разными способами, или, как говорят, по разным каналам. Для простоты ограничимся рассмотрением двух каналов: один предполагает обмен энергией через *передачу теплоты,* а другой — через *совершение работы. Первое начало термодинамики* есть не что иное, как закон сохранения энергии с учетом возможности обмена энергией между объектом и средой по разным каналам:

$$\Delta U = A + Q, \tag{4.1}$$

где $\Delta U = U_2 - U_1$ — приращение внутренней энергии объекта (U_1 и U_2 — энергии начального и конечного состояний объекта), A — работа, совершенная внешними телами над объектом, Q — теплота, переданная объекту. Заметим, что в отличие от внутренней энергии, являющейся *функцией состояния* объекта (ее величина изменяется по мере перехода объекта из одних состояний в другие), ни работа, ни теплота функциями состояния объекта не являются. Говорить, что в таком-то состоянии у тела столько-то теплоты, так же бессмысленно, как говорить, что у тела столько-то работы. Теплота Q и работа A в формуле (4.1) — это осуществляемые по разным каналам изменения энергии объекта. Будем рассматривать наиболее простую макросистему — *идеальный газ* (m — масса газа). Внутренняя энергия идеального газа пропорциональна абсолютной температуре газа T и не зависит от

объема V, занимаемого газом. При помощи поршня будем менять объем газа. Вдвигая поршень и сжимая газ, мы тем самым совершаем некоторую работу A над газом. При расширении газ совершает работу A' над поршнем: $A' = -A$. Работа связана с изменением объема газа. Численно она равна площади под графиком зависимости $p(V)$, описывающей рассматриваемый процесс (p — давление газа); площадь под графиком вычисляется на участке от $V = V_1$ до $V = V_2$, где V_1 и V_2 — начальный и конечный объемы газа.

В качестве примера рассмотрим с точки зрения первого начала термодинамики два процесса расширения газа — *изотермический* и *адиабатический*. Первый процесс происходит при неизменной температуре газа, а второй в условиях, когда нет теплообмена между газом и окружающей средой. Изменение объема газа должно совершаться достаточно медленно (по сравнению с процессами установления теплового равновесия в газе), тогда в любой момент времени состояние газа может рассматриваться как термодинамически равновесное. Иначе говоря, мы полагаем, что переход газа из одного термодинамически равновесного состояния в другое происходит как бы через последовательность промежуточных равновесных состояний.

При *изотермическом* расширении температура газа постоянна и, следовательно, $\Delta U = 0 (U_1 = U_2)$. Учитывая это, получаем из (4.1):

$$-A = Q, \text{ или } A' = Q. \tag{4.2}$$

Расширяющийся газ совершает работу, равную теплоте, получаемой им от окружающих тел в процессе расширения. При *адиабатическом* расширении нет теплообмена со средой ($Q = 0$). Следовательно,

$$\Delta U = A, \text{ или } A' = -\Delta U. \tag{4.3}$$

Расширяющийся газ совершает работу за счет уменьшения внутренней энергии, температура газа при этом убывает. На рисунке 4.1 условным образом показаны оба рассматриваемых процесса. Там же эти процессы изображены графически в осях (V, p). Найдем работу A', совершаемую газом при изотермическом расширении от объема $V = V_1$ до $V = V_2$. Работа численно равна площади, под соответствующим графиком $p(V)$, показанной на рисунке желтым цветом:

$$A' = \int_{V_1}^{V_2} p(V)\, dV. \tag{4.4}$$

Используя уравнение состояния идеального газа (уравнение Менделеева — Клапейрона), представим

$$p = mRT/MV, \tag{4.5}$$

где M — молярная масса газа, R — универсальная газовая постоянная. Подставляя (4.5) в (4.4) и учитывая при этом, что

температура газа постоянна, получаем

$$A = \frac{mRT}{M} \int_{V_1}^{V_2} \frac{1}{V} dV = \frac{mRT}{M} \ln \frac{V_2}{V_1} \quad (4.6)$$

(символом ln обозначается логарифм по основанию $e = 2{,}71828...$).

Цикл Карно. В 1824 году в Париже вышла книга 28-летнего инженера Сади Карно «Размышления о движущей силе огня и о машинах, способных развивать эту силу». К сожалению, глубина мыслей, изложенных в этой книге, была оценена лишь много лет спустя, уже после смерти ее автора. Карно исследовал вопросы, связанные с получением работы в *тепловых машинах*. Он показал, что для создания тепловой машины недостаточно иметь только нагретое тело, требуется еще и второе тело, имеющее более низкую температуру. Первое тело условно называют нагревателем, а второе холодильником. Кроме нагревателя и холодильника, необходимо рабочее тело (жидкость, пар, газ), которое передает теплоту от нагревателя к холодильнику и попутно совершает полезную работу.

Выбрав в качестве рабочего тела идеальный газ, Карно рассмотрел замкнутый цикл, состоящий из двух изотерм и двух адиабат. Впоследствии этот цикл стали называть *циклом Карно*. Он показан на рисунке 4.2. Пусть T_1 — температура нагревателя, а T_2 — холодильника. На участке *1 — 2* (изотерма при T_1) газ получает от нагревателя теплоту Q_1 и, расширяясь, расходует ее на работу A'_1. На участке *2 — 3* (адиабата) газ совершает работу A'_3, при этом его температура падает и становится равной T_2. На участке *3—4* (изотерма при T_2) газ отдает холодильнику теплоту Q_2, равную работе A_2 по сжатию газа. На участке *4—1* (адиабата) работа A_4 по сжатию газа переходит во внутреннюю энергию, температура газа повышается до T_1. В результате рабочее тело возвращается в исходное состояние *1*. Предположим, что тепловая машина работает по циклу Карно. Газ получает от нагревателя теплоту Q_1 и отдает холодильнику теплоту Q_2. В соответствии с (4.2) запишем: $Q_1 = A'_1$, $|Q_2| = A_2$. Здесь мы учитываем, что $Q > 0$, когда теплота передается газу, и $Q < 0$, когда теплота отбирается от газа. Из рисунка 4.2 видно, что площадь под изотермой *3 — 4* меньше, чем под изотермой *1 — 2*, следовательно, $A_2 {}^{-} A'_1$. Таким образом, $|Q_2| < Q_1$, т. е. газ отдает холодильнику теплоты меньше, чем получает ее от нагревателя. В то же время внутренняя энергия газа за цикл остается прежней. Следовательно, разность $Q_1 - |Q_2|$ равна полезной работе, совершенной тепловой машиной за цикл. Отсюда находим КПД тепловой машины:

$$\eta = (Q_1 - |Q_2|)/Q_1. \quad (4.7)$$

Карно показал, что

$$Q_1/T_1 = |Q_2|/T_2. \quad (4.8)$$

Это позволяет переписать выражение (4.7) в виде:

$$\eta = (T_1 - T_2)/T_1. \quad (4.9)$$

Определяемый формулами (4.7) и (4.9) КПД тепловой машины есть наибольший возможный КПД. В реальных тепловых машинах КПД всегда меньше — из-за неизбежных необратимых процессов.

Обратимые и необратимые процессы. Понятия обратимого и необратимого процессов очень важны для термодинамики. Процесс называется *обратимым*, если система (рабочее тело) все время находится в тепловом равновесии, непрерывно переходя из одних равновесных состояний в другие. Такой процесс полностью (на протяжении всего хода) управляется изменением тех или иных параметров, например температуры или объема. Изменяя параметры в обратном направлении, можно в точности обратить рассматриваемый процесс. Обратимые процессы называют также *равновесными*.

Известные законы Бойля — Мариотта и Гей-Люссака описывают обратимые процессы в идеальном газе. Полученные выше выражения (4.7) и (4.9) относятся к обратимому циклу Карно, его называют также идеальным циклом Карно. Каждый участок цикла, а также весь цикл в целом можно при желании в точности обратить.

Необратимыми называются процессы, ходом которых нельзя управлять. Они протекают самостоятельно, иначе говоря, самопроизвольно. Как следствие, нельзя обратить ход таких процессов. Ранее отмечалось, что, будучи выведена из теплового равновесия, система самопроизвольно переходит в термодинамически равновесное состояние. Процессы, связанные с переходом системы из неравновесного состояния в равновесное,— это необратимые процессы. Их называют также *неравновесными* процессами.

Вот некоторые примеры необратимых процессов: переход теплоты от нагретого тела к холодному; взаимное перемешивание двух различных газов, попавших в один объем; расширение газа в пустоту. Все эти процессы протекают самопроизвольно, без управления извне. В то же время не наблюдается самопроизвольный переход теплоты от холодного тела к нагретому. Не происходит самопроизвольное разделение газовых компонентов из смеси газов. Не было случаев, чтобы газ вдруг самопроизвольно сжался. Следует подчеркнуть: *всякий необратимый процесс характеризуется определенной направленностью.* Он развивается в каком-то одном направлении и не развивается в обратном. Какое направление развития процесса оказывается дозволенным, а какое запрещенным,— этими вопросами ведает второе начало термодинамики.

Второе начало термодинамики. Одну из первых формулировок *второго начала термодинамики* дал известный английский физик Уильям Томсон (лорд Кельвин): «Невозможно построить периодически действующую машину, вся деятельность которой сводилась бы к поднятию некоторого груза и соответствующему охлаждению теплового резервуара». Это означает, что нельзя создать

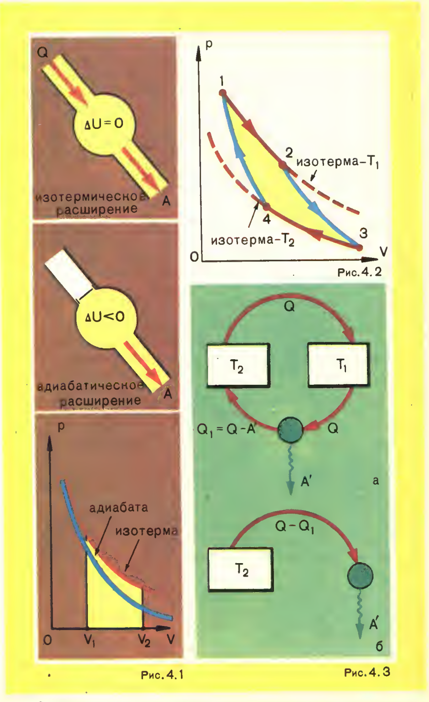


Рис. 4.1

Рис. 4.2

Рис. 4.3

машину, которая производила бы полезную работу просто за счет некоторого уменьшения внутренней энергии какой-нибудь среды, например за счет уменьшения внутренней энергии воды в море. Такую машину Кельвин назвал *вечным двигателем второго рода*. Если в прежних проектах вечных двигателей нарушался закон сохранения энергии (вечные двигатели первого рода), то здесь такого нарушения нет. Вечный двигатель второго рода не противоречит первому началу термодинамики, его запрещает второе начало.

В 1850 году немецкий физик Р. Клаузиус сформулировал второе начало термодинамики так: «Переход теплоты от более холодного тела к более теплому не может совершаться без компенсации». Полезно продемонстрировать *эквивалентность* формулировок Кельвина и Клаузиуса. Если бы нам удалось, вопреки формулировке Кельвина, «извлечь» теплоту из некоторой среды и при помощи циклического процесса обратить ее в полезную работу, то затем, используя трение, мы смогли бы превратить эту работу в теплоту при более высокой температуре. Это противоречило бы формулировке Клаузиуса, так как означало бы перенос теплоты от менее нагретого тела к более нагретому при помощи замкнутого цикла без совершения какой-либо работы внешними телами.

Предположим, с другой стороны, что, вопреки формулировке Клаузиуса, нам удалось перенести некоторое количество теплоты Q от холодного тела (температура T_2) к нагретому (T_1). При последующем естественном переходе этой теплоты от нагретого тела к холодному можно получить некоторую полезную работу A', а остаток теплоты $Q_1 = Q - A'$ возвратится к холодному телу. Этот процесс схематически показан на рисунке 4.3, *а*. Ясно, что такой процесс соответствует прямому превращению теплоты $Q - Q_1$ в работу A' (рис. 4.3, *б*), что, очевидно, противоречит формулировке Кельвина.

Энтропия. Внимательно знакомясь с исследованиями, выполненными Карно, Клаузиус обратил внимание на то, что соотношение (4.8) похоже на закон сохранения. Величина Q_1/T_1, «отобранная» рабочим телом у нагревателя, равна величине $|Q_2|/T_2$, «переданной» затем холодильнику. Клаузиус постулировал существование некоторой величины S, являющейся, подобно внутренней энергии, *функцией состояния* тела. Если к рабочему телу (в данном случае к идеальному газу) подводится теплота Q при температуре T, то величина S получает приращение

$$\Delta S = Q/T. \qquad (4.10)$$

Клаузиус назвал величину S *энтропией*.

На участке *1—2* цикла Карно (см. рис. 4.2) происходит передача теплоты Q_1 от нагревателя к рабочему телу при температуре T_1, при этом энтропия рабочего тела возрастает на $\Delta S_1 = Q_1/T_1$. На участках *2 — 3* и *4 — 1* передачи теплоты нет, поэтому энтропия рабочего тела не изменяется. На участке *3 — 4* происходит передача теплоты Q_2 от рабочего тела к холодильнику при темпера-

туре T_2, при этом энтропия тела уменьшается на $|\Delta S_2|=|Q_2|/T_2$ ($\Delta S_2<0$). Согласно (4.8) и (4.10),

$$\Delta S_1+\Delta S_2=0. \qquad (4.11)$$

Таким образом, по окончании идеального (обратимого) цикла Карно энтропия рабочего тела принимает исходное значение.

Заметим, что энтропия может быть определена, как такая функция состояния тела (системы), которая сохраняется при адиабатических процессах. Аналогичным образом температуру можно рассматривать как функцию состояния системы, сохраняющуюся при изотермических процессах.

В дальнейшем нам понадобится свойство энтропии, называемое *аддитивностью*. Это свойство означает, что *энтропия системы есть сумма энтропий частей системы*. Свойством аддитивности обладают также внутренняя энергия, масса, объем. Однако ни температура, ни давление таким свойством не обладают.

Второе начало термодинамики как закон возрастания энтропии замкнутой системы при необратимых процессах. Используя понятие энтропии, можно дать следующую формулировку второго начала термодинамики: *любой необратимый процесс в замкнутой системе идет таким образом, чтобы энтропия системы при этом возрастала.* Рассмотрим в качестве примера следующий необратимый процесс. Пусть некая замкнутая система состоит из подсистем *1* и *2*, имеющих температуры T_1 и T_2 соответственно. Предположим, что небольшое количество теплоты ΔQ переходит от подсистемы *1* к подсистеме *2*, при этом температуры подсистем практически не изменяются. Энтропия подсистемы *1* уменьшается на $\Delta Q/T_1$ ($S_1=-\Delta Q/T_1$), а энтропия подсистемы *2* увеличивается на $\Delta S_2=\Delta Q/T_2$. Энтропия системы есть сумма энтропий ее подсистем, следовательно, приращение энтропии системы составит

$$\Delta S=\Delta S_1+\Delta S_2=\Delta Q\left(\frac{1}{T_2}-\frac{1}{T_1}\right). \qquad (4.12)$$

Процесс перехода теплоты от подсистемы *1* к подсистеме *2* есть необратимый процесс, он совершается, если $T_1>T_2$. Учитывая это неравенство, заключаем из (4.12), что $\Delta S>0$. Итак, мы убедились, что процесс передачи теплоты от нагретого тела к холодному сопровождается повышением энтропии системы, состоящей из рассматриваемых тел.

Возрастание энтропии при необратимых процессах является обязательным законом лишь для *замкнутых* систем. Если же система незамкнутая, то возможно уменьшение ее энтропии. Так, если какие-то внешние тела совершают над системой работу, то можно реализовать передачу некоторого количества теплоты от холодильника к нагревателю. Существенно, что если при этом включить в систему, наряду с нагревателем, холодильником и рабочим телом, также все те тела, которые совершали работу (т. е. если снова перейти к рассмотрению замкнутой системы), то суммарная энтропия такой системы будет возрастать.

Сформулируем основные выводы, касающиеся изменения энтропии системы.

Первый вывод. Если система *замкнутая*, то ее энтропия со временем не убывает:

$$\Delta S \geqslant 0. \tag{4.13}$$

При обратимых процессах в системе энтропия системы остается неизменной. При необратимых процессах энтропия системы возрастает. Возрастание энтропии можно рассматривать как меру необратимости процессов, происходящих в системе.

Второй вывод. Об изменении энтропии *незамкнутой* системы ничего определенного в общем случае сказать нельзя. Она может оставаться постоянной, возрастать и даже уменьшаться.

Загадки термодинамики. Эти загадки концентрируются вокруг второго начала термодинамики. Диктуя определенную *направленность* процессов в природе, это начало вводит принципиальную необратимость. Как объяснить эту необратимость с точки зрения физики? Почему теплота переходит от нагретого тела к холодному, а обратно сама по себе не переходит? Почему газ расширяется в пустоту, но не сжимается самопроизвольно? Почему, оказавшись в одном объеме, разные газы перемешиваются, а вот самопроизвольного разделения газов из смеси не происходит? Молот ударяет по наковальне. Происходит некоторое повышение температуры наковальни. Почему, сколько бы мы ни нагревали наковальню с лежащим на ней молотом, нельзя добиться обратного эффекта — отскакивания молота от наковальни?

Подобных «почему» можно было бы предложить очень много. Термодинамика принципиально не отвечает на все эти «почему». Ответ надо искать в молекулярно-кинетической теории вещества. Следует обратиться к картине хаотически движущихся молекул.

Молекулы в газе и вероятность

Диалог с участием автора. Вообразим, что мы беседуем с физиком, жившим в 60-х годах прошлого столетия. Нам не понадобится «машина времени». Просто мы будем полагать, что наш собеседник придерживается взглядов, характерных для физиков середины XIX века. Тех самых физиков, многие из которых позднее в 70-х годах, не смогли понять и принять идеи великого австрийского физика Людвига Больцмана (1844—1906).

Итак, перенесемся мысленно в 1861 год.

Автор. Будем рассматривать газ как коллектив из большого числа хаотически движущихся молекул.

Собеседник. Не возражаю. Мне знакомы недавние исследования Джеймса Клерка Максвелла, вычислившего распределение молекул в газе по скоростям.

Автор. Мне хотелось бы поговорить не о распределении, установленном Максвеллом, а о некоторых вопросах принципиального характера. Дело в том, что переход от термодинамического рас-

смотрения к рассмотрению движений молекул есть качественный скачок. Это есть переход от *динамических* законов с их жестко детерминированными зависимостями к *вероятностным* законам, которые управляют процессами в больших коллективах молекул.

Собеседник. Однако движениями молекул управляют вовсе не законы вероятности, а законы классической динамики Ньютона. Допустим, что в какой-то момент времени мы задали координаты и скорости всех молекул в газе. Допустим, что мы смогли проследить за всеми соударениями молекул между собой и со стенками сосуда. Ясно, что в таком случае мы могли бы совершенно точно предсказать, где окажется в некоторый момент времени та или иная молекула и какую скорость она будет при этом иметь.

Автор. Вас не смущает, что при этом вы уподобляетесь тому самому сверхсуществу, о котором писал в свое время Лаплас?

Собеседник. Передо мной конкретная задача из области механики. Правда, с чрезвычайно большим количеством тел.

Автор. В кубическом сантиметре газа при нормальных условиях находится около 10^{19} молекул. Вам пришлось бы иметь дело с задачей, где надо учесть взаимодействия около 10^{20} тел.

Собеседник. Трудности, конечно, исключительно велики. Но они имеют не принципиальный, а чисто технический характер. И поскольку наши вычислительные возможности весьма ограниченны, приходится поневоле прибегать к вероятностям — вероятности молекуле попасть в такой-то объем, вероятности ей иметь скорость в таком-то интервале значений и т. д.

Автор. Итак, вы полагаете, что использование вероятностей связано лишь с практической невыполнимостью чересчур громоздких расчетов, тогда как коллектив молекул в принципе ведет себя в соответствии с законами Ньютона, которым подчиняются отдельные молекулы.

Собеседник. Именно так. Поэтому и не усматриваю упомянутого вами ранее качественного скачка.

Автор. У меня есть, по крайней мере, три веских аргумента в пользу того, что вероятностное описание больших коллективов молекул необходимо принципиально, что случайность присутствует в самой природе этих коллективов, а не связана, как вы полагаете, лишь с неполнотой наших знаний и неумением выполнять громоздкие расчеты.

Собеседник. Хотелось бы познакомиться с этими аргументами.

Автор. Начну с первого. Допустим, что существует, как вы утверждаете, жесткая система строго детерминированных связей между молекулами в газе. Представим, что в такой системе вдруг произойдет исчезновение какого-то числа молекул (молекулы попросту вылетели из сосуда через какую-нибудь щель). Ясно, что вместе с этими молекулами исчезнут и все предопределенные их присутствием в газе последующие столкновения с другими молекулами, что, в свою очередь, изменит поведение этих других молекул. Все это не может не сказаться на всей системе жестких взаимосвязей и, как следствие, на поведении коллектива молекул в целом. Однако

известно, что можно совершенно безболезненно с точки зрения газа в целом изъять вдруг большое количество молекул (например, 10^{12} молекул и даже существенно больше). При этом свойства газа, его поведение нисколько не изменятся. Разве это не указывает на то, что динамические закономерности, управляющие поведением отдельных молекул, не имеют фактически какого-либо отношения к поведению газа?

Собеседник. И все же трудно поверить, что молекулы подчиняются одним законам, а коллектив из этих молекул — совсем другим.

Автор. Но это именно так. И мой второй аргумент еще более подчеркнет это принципиальное обстоятельство. Приведу простые примеры. Камень бросают из точки *A* под углом к горизонту (рис. 4.4, *а*). Допустим, что в точке *B* траектории полета мы мысленно изменили направление скорости камня на противоположное. Ясно, что камень возвратится в точку *A* и будет иметь в ней ту же скорость (по модулю), какую имел в момент бросания. Получается, что летящий камень как бы «помнит» свою историю.

Собеседник. Это естественно. Ведь каждое мгновенное состояние летящего камня предопределено предыдущим и, в свою очередь, определяет последующее.

Автор. Другой пример: шар упруго ударяется о некоторую стенку и отскакивает (рис. 4.4, *б*). Если в точке *B* изменить направление скорости шара на противоположное, ситуация повторится в обратном порядке — шар ударится о стенку и возвратится в точку *A*.

Эти примеры я привел для того, чтобы проиллюстрировать важную мысль — для движений, определяемых законами классической механики, характерна своего рода «память» о прошлом. Недаром эти движения обратимы.

Иное дело — поведение газа. Представьте себе следующую ситуацию. Имеется пучок молекул, скорости которых направлены строго параллельно друг другу. Попадая внутрь некоего сосуда, молекулы испытывают множество столкновений друг с другом и со стенками. В результате газ из молекул приходит в состояние термодинамического равновесия. В этом состоянии утрачивается какая-либо «память» о прошлом. Можно сказать, что пришедший в тепловое равновесие газ как бы «забыл» свою предысторию, не помнит, каким именно образом он пришел в состояние равновесия. Поэтому не имеет смысла говорить о том, чтобы обратить всю ситуацию,— молекулы газа не соберутся в единый пучок, выходящий из сосуда в строго определенном направлении.

Примеров подобной «забывчивости» можно привести много. Предположим, что по одну сторону перегородки в сосуде находится один газ, а по другую сторону другой. Если убрать перегородку, произойдет перемешивание молекул обоих газов. Очевидно, не следует ожидать обращения этой картины — молекулы не будут распределяться по «своим» половинам сосуда. Можно сказать, что газ,

в котором произошло перемешивание, уже «не помнит» предыстории.

Собеседник. Вы хотите сказать, что данное равновесное состояние газа не предопределено предыдущими состояниями газа?

Автор. Когда употребляют слово «предопределено», имеют в виду жесткую, однозначную предопределенность. Такой здесь действительно нет. Данное равновесное состояние газа можно реализовать, исходя из различных начальных состояний. Ни об одном из них вы не найдете никакой информации, исследуя газ в тепловом равновесии. Это и означает, что «газ забыл свою предысторию».

Собеседник. В этом я с вами согласен.

Автор. А когда происходит «утрата предыстории»? Она происходит тогда, когда на сцене появляется *случайность*. Вы бросаете игральный кубик — выпадает, допустим, четверка. Вы снова бросаете — выпадает единица. Выпадение единицы никак не связано с тем, что перед этим выпала четверка. Вы выполняете множество бросаний и получаете набор цифр. Этот набор обнаруживает устойчивость (например, четверка встречается примерно в одной шестой всех случаев). Указанная устойчивость не имеет предыстории — она не связана с выпадением тех или иных цифр при отдельных бросаниях кубика.

Точно так же и в случае с газом. Утрата предыстории показывает, что здесь мы имеем дело с закономерностями *статистического типа,* закономерностями, где случайность играет принципиальную роль.

Собеседник. Казалось бы, все было так ясно. Была создана механика Ньютона. Потом появились температура и давление газа. Использовав представления о молекулах, мы свели эти физические величины к механическим, связав температуру с энергией молекул, а давление газа с импульсом, передаваемым стенке со стороны ударяющихся об нее молекул. Таким образом, законы механики были и продолжали оставаться фундаментальными законами. Вы же предлагаете поставить наравне с законами механики также вероятностные законы.

Автор. Полагаю, вам известно, что не все термодинамические величины имеют свои аналоги в классической механике. И вот вам мой третий аргумент — у энтропии нет механического аналога. Одного лишь существования такой величины, как энтропия, достаточно, чтобы опровергнуть тезис об исчерпывающей фундаментальности законов классической механики.

Собеседник. Об энтропии мне вообще не хотелось бы говорить...

Окончим на этом слегка затянувшийся диалог. По условию, он относился к 1861 году. Поэтому автор не стал обращаться к аргументам, которые не могли быть известны в то время. Если бы не эти соображения, автор мог бы привести еще два аргумента в пользу своей позиции. Во-первых, он указал бы на то, что энтропия принципиально выражается через вероятность и что именно это обстоятельство позволяет дать объяснение всем загадкам термо-

динамики. Подробнее мы поговорим об этом в последующих параграфах данной главы. Во-вторых, как это следует из квантовой физики, допущение нашего собеседника о том, что он в принципе может задать одновременно координаты и скорости молекул, оказывается несостоятельным. Этого нельзя сделать по принципиальным соображениям, о чем мы подробнее поговорим позднее, в главе пятой.

А теперь обратимся к картине движущихся в газе молекул.

Картина движения молекул в газе, находящемся в термодинамическом равновесии. Пусть масса газа m находится в тепловом равновесии. Газ занимает объем V и характеризуется температурой T и давлением p.

Каждая молекула газа движется с постоянной по модулю и направлению скоростью, пока не столкнется с какой-либо другой молекулой или не налетит на стенку. В целом картина движения молекул представляется хаотичной — молекулы движутся в разных направлениях с разными скоростями, происходят беспорядочные столкновения, приводящие к изменениям направления движения и значения скорости молекул. Сделаем мысленную «фотографию» положений молекул в некоторый момент времени. Она выглядит примерно так, как показано на рисунке 4.5, где для простоты рассматриваются не три, а только два измерения (картина плоская). Видно, что точки-молекулы достаточно равномерно заполняют объем сосуда (границы объема показаны на рисунке в виде квадрата). Пусть N — полное число молекул в сосуде; $N = N_A m/M$, где N_A — постоянная Авогадро. В любом месте внутри сосуда в любой момент времени число молекул в единице объема будет в среднем одним и тем же: N/V. Та или иная молекула с одинаковой вероятностью может быть обнаружена в любом месте внутри сосуда.

Обозначим через $G(x, y, z)\,\Delta x \Delta y \Delta z$ вероятность обнаружить молекулу внутри объема $\Delta V = \Delta x \Delta y \Delta z$ у точки с координатами x, y, z. Точнее говоря, это есть вероятность того, что x-координата молекулы окажется в интервале значений от x до $x+\Delta x$, y-координата — от y до $y+\Delta y$, z-координата — от z до $z+\Delta z$. При достаточно малых Δx, Δy, Δz функция $G(x, y, z)$ может рассматриваться как *плотность вероятности* обнаружить молекулу в точке (x, y, z). Плотность вероятности в данном случае не зависит от координат: $G=\text{const}$. Учитывая, что вероятность обнаружить молекулу где-нибудь внутри сосуда равна единице, запишем: $\int_V G\,dV = 1$, или $G\int_V dV = GV = 1$. Таким образом, $G = 1/V$.

Итак, где бы внутри сосуда ни был выбран единичный объем, вероятность того, что некоторая молекула окажется внутри этого объема, равна $1/V$, т. е. равна отношению единичного объема к объему сосуда. Обобщая этот вывод, можем утверждать, что вероятность обнаружить некоторую молекулу внутри объема V_0 равна V_0/V.

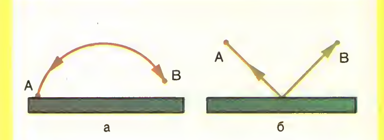


Рис. 4.4

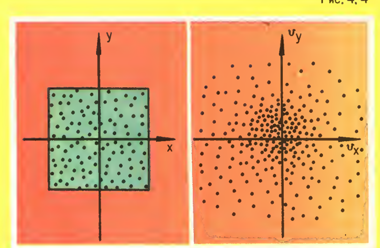


Рис. 4.5

Рис. 4.6

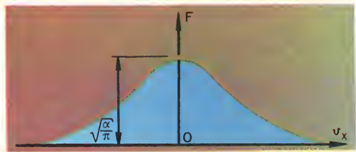


Рис. 4.7

Теперь перейдем к обсуждению скоростей молекул в газе. Заведомо ясно, что говорить о равновероятности различных значений скорости не приходится: число молекул должно уменьшаться как в области очень малых, так и в области очень высоких скоростей. Рассматривая скорости молекул, удобно использовать *пространство скоростей,* в котором по осям координат откладываются значения проекций скорости молекулы (v_x, v_y, v_z). На рисунке 4.6 показаны для простоты только две оси — ось v_x и ось v_y (двумерное пространство скоростей). Рисунок показывает зафиксированную в некоторый момент времени картину распределения скоростей молекул в газе. Каждая точка на рисунке соответствует молекуле. Абсцисса точки есть x-проекция скорости молекулы, а ордината — y-проекция.

Интересно сопоставить рисунки 4.5 и 4.6. На первом точки располагаются внутри некоторой области, причем достаточно равномерно. На втором разброс точек в принципе не ограничен какими-либо рамками. Точки явно концентрируются вблизи начала координат. Это означает, что проекция скорости молекулы может быть в принципе сколь угодно большой, однако наиболее вероятны проекция скорости вблизи нуля. Картина разброса точек на рисунке 4.6 симметрична относительно поворота вокруг начала координат. Это означает, что равновероятны все направления движения: ту или иную молекулу с одинаковой вероятностью можно обнаружить движущейся в любом направлении.

Чтобы получить представление о картине движения молекул в газе, надо использовать оба рисунка. А еще лучше было бы вместо каждого из рисунков рассмотреть последовательность своеобразных кинокадров, снятых для ряда последовательных моментов времени. Тогда бы мы увидели, что точки на рисунке 4.5 движутся в разных направлениях — в моменты столкновений их траектории претерпевают изломы. На рисунке 4.6 точки не движутся. Зато они то тут, то там вдруг исчезают или, наоборот, появляются. Всякий раз исчезает одновременно какая-то пара точек и тут же возникают какие-то две новые точки — это есть результат столкновения каких-то двух молекул.

Закон распределения Максвелла. Пусть $F(v_x)\Delta v_x$ — вероятность того, что некоторая молекула (в некоторый момент времени) имеет x-проекцию скорости в интервале значений от v_x до $v_x+\Delta v_x$. При этом остальные две проекции скорости молекулы могут быть какими угодно. При малых Δv_x функция $F(v_x)$ есть плотность вероятности обнаружить молекулу, имеющую проекцию скорости v_x.

Великий английский физик Джеймс Клерк Максвелл (1831—1879) показал, что плотность вероятности $F(v_x)$ соответствует *закону Гаусса:*

$$F(v_x) = Ae^{-\alpha v_x^2}, \qquad (4.14)$$

где α — некоторый параметр ($\alpha > 0$). Постоянная A определяется

нз равенства

$$\int_{-\infty}^{\infty} F(v_x)dv_x=1, \qquad (4.15)$$

выражающего тот факт, что вероятность молекуле иметь какую-ннбудь x-проекцню скорости равна еднннце. Подставляя (4.14) в (4.15), получаем $A\int_{-\infty}^{\infty} e^{-\alpha v_x^2}\,dv_x=1$. Интеграл $\int_{-\infty}^{\infty} e^{-\alpha v_x^2}\,dv_x$ нзвестен в математнке как *интеграл Пуассона,* он равен $\sqrt{\pi/\alpha}$. Следовательно, $A=\sqrt{\alpha/\pi}$. В нтоге перепншем соотношение (4.14) в внде:

$$F(v_x)=\sqrt{\frac{\alpha}{\pi}}\,e^{-\alpha v_x^2}. \qquad (4.16)$$

Такнми же функциямн F опнсываются плотностн вероятностн для y- и z-проекцнй скоростн молекулы. Графнк функции $F(v_x)$ показан на рнсунке 4.7. Пусть $f(v_x, v_y, v_z)$ — плотность вероятности обнаружнть молекулу с проекцнямн скоростн v_x, v_y, v_z. Используя теорему умножения вероятностей, представнм:

$$f(v_x,\ v_y,\ v_z)\Delta v_x\Delta v_y\Delta v_z=[F(v_x)\Delta v_x][F(v_y)\Delta v_y][F(v_z)\Delta v_z].$$

Следовательно,

$$f(v_x, v_y, v_z)=\left(\frac{\alpha}{\pi}\right)^{3/2}e^{-\alpha(v_x^2+v_y^2+v_z^2)}=\left(\frac{\alpha}{\pi}\right)^{3/2}e^{-\alpha v^2}. \qquad (4.17)$$

Мы вндим, что завнснмость плотностн вероятности от проекцнй скоростн выглядит как завнснмость от $v_x^2+v_y^2+v_z^2=v^2$. Этого следовало ожндать, поскольку, как уже отмечалось, различные направлення скоростн равновероятны, так что плотность вероятностн может завнсеть лншь от модуля скоростн молекулы.

Итак, вероятность обнаружить молекулу с проекцнями скоростн в ннтервалах $v_x - v_x+\Delta v_x$, $v_y - v_y+\Delta v_y$, $v_z - v_z+\Delta v_z$ есть

$$\Delta w_{\vec{v}}=\left(\frac{\alpha}{\pi}\right)^{3/2}e^{-\alpha v^2}\Delta v_x\Delta v_y\Delta v_z, \qquad (4.18)$$

где $v^2=v_x^2+v_y^2+v_z^2$.

Сделаем еще однн шаг — учнтывая сказанное выше о равновероятностн направленнй скорости, будем рассматрнвать вероятность обнаружить молекулу, у которой модуль скоростн находнтся в ннтервале от v до $v+\Delta v$. Направленне скоростн молекулы может быть какнм угодно. Рассматрнвается только модуль скоростн. В связн с этнм обратнмся к пространству скоростей (рнс. 4.8). Рассматрнвавшаяся выше вероятность $\Delta w_{\vec{v}}$ (см. 4.18) есть вероятность обнаружнть молекулу в «объеме» ΔV_v, показанном на рнсунке 4.8, *а* (слово «объем» взято в кавычкн, чтобы напомнить,

что речь идет не об обычном пространстве, а о пространстве скоростей). Теперь же мы хотим рассматривать вероятность обнаружения молекулы внутри показанного на рисунке 4.8, *б* шарового слоя, заключенного между сферами с радиусами v и $v+\Delta v$. «Объем» этого слоя равен произведению площади сферы радиуса v на толщину слоя Δv — это есть $4\pi v^2 \Delta v$. Таким образом, искомая вероятность имеет вид:

$$\Delta w_v = \left(\frac{\alpha}{\pi}\right)^{3/2} e^{-\alpha v^2} 4\pi v^2 \Delta v. \tag{4.19}$$

Эта формула выражает закон распределения молекул идеального газа по модулю скорости — *закон распределения Максвелла*. Плотность вероятности $g(v)=\Delta w_v/\Delta v$ показана на рисунке 4.9. Она стремится к нулю как при $v\to 0$, так и при $v\to\infty$. При $v\to 0$ обращается в нуль «объем» изображенного на рисунке 4.8, *б* шарового слоя, а при $v\to\infty$ стремится к нулю множитель $e^{-\alpha v^2}$ в законе распределения.

Случайность и необходимость в картине движущихся молекул. Предположим, что нам удалось бы для некоторого момента времени зафиксировать положения и скорости всех молекул в рассматриваемом объеме газа. Мысленно разобьем весь объем на множество одинаковых ячеек и будем просматривать нашу мгновенную «фотографию», переходя от одних ячеек к другим. При этом окажется, что число молекул изменяется от ячейки к ячейке случайным образом. Выделим только те молекулы, модуль скорости которых попадает в некоторый интервал от v до $v+\Delta v$. Число таких молекул изменяется от ячейки к ячейке случайным образом. Разобьем полный телесный угол 4π стерадиан на одинаковые телесные углы, ориентированные в разных направлениях. Число молекул, направление скорости которых попадает в тот или иной «элементарный» телесный угол, изменяется от одного такого угла к другому случайным образом.

Можно было бы поступить иначе — сосредоточить внимание на какой-либо одной ячейке или на каком-либо одном «элементарном» телесном угле, но делать «фотографии» в разные моменты времени. Числа молекул (в ячейке или в телесном угле), рассматриваемые в разные моменты времени, также обнаруживают случайные изменения от одного момента времени к другому.

Желая подчеркнуть фактор *случайного* в картине движущихся молекул, применяют термин «хаотический»: хаотические столкновения молекул друг с другом, хаотически ориентированные направления скоростей молекул и вообще — хаотическое тепловое движение молекул. В этой картине «хаоса» просматривается, однако, *порядок*, или, иначе говоря, *необходимость*, то, что мы не раз уже называли *статистической устойчивостью*.

Статистическая устойчивость проявляется в существовании совершенно определенных вероятностей: вероятности молекуле ока-

заться в объеме ΔV (эта вероятность равна $\Delta V/V$); вероятности молекуле двигаться в пределах телесного угла $\Delta\Omega$ (эта вероятность равна $\Delta\Omega/4\pi$); вероятности молекуле иметь модуль скорости в интервале от v до $v+\Delta v$ (эта вероятность описывается формулой 4.19).

Число молекул в единице объема, имеющих модуль скорости в интервале от v до $v+\Delta v$, с высокой степенью точности равно

$$\Delta n=\frac{N}{V}\Delta w_v=4\pi\frac{N}{V}\left(\frac{\alpha}{\pi}\right)^{3/2}e^{-\alpha v^2}v^2\Delta v. \tag{4.20}$$

Столкновения молекул приводят к тому, что часть из них уходит из рассматриваемого интервала значений скорости; однако другие столкновения приводят к появлению новых молекул в указанном интервале. В результате поддерживается определенный порядок — число молекул в данном интервале значений скорости остается практически неизменным и определяется соотношением (4.20). Подчеркнем, что случайность и необходимость выступают, как всегда, в диалектическом единстве. Столкновения большого числа молекул обусловливают случайность в картине движущихся молекул. Они же поддерживают термодинамически равновесное состояние газа, характеризующееся определенными вероятностями, через которые и проявляется статистическая устойчивость.

Давление и температура идеального газа

Давление как результат «молекулярной бомбардировки». Стенки сосуда, внутри которого находится газ, испытывают множество ударов со стороны молекул газа. Эта «молекулярная бомбардировка» и приводит к тому, что газ оказывает на стенку давление. Выберем ось x перпендикулярно к стенке. Из рисунка 4.10, *а* видно, что x-проекция импульса молекулы изменяется при упругом ударе о стенку на $2m_0v_x$, где m_0 — масса молекулы. Это означает, что, ударяясь о стенку, молекула передает ей импульс, равный $2m_0v_x$. Сначала будем учитывать лишь те молекулы газа, x-проекция скорости которых находится в интервале от v_x до $v_x+\Delta v_x$ (заметим при этом, что $v_x>0$, иначе молекула будет лететь не к стенке, а от нее); другие проекции скорости молекулы могут быть какими угодно. Число ударов со стороны рассматриваемых молекул, испытываемое участком стенки площадью s в единицу времени, равно числу таких молекул в объеме, равном sv_x (рис. 4.10, *б*). (Читателя не должно смущать то, что произведение sv_x не имеет размерности объема. В действительности мы имеем здесь дело с произведением $s(\text{см}^2)\times v_x(\text{см/с})\times 1\,(\text{с})$.) С учетом (4.16) это число ударов есть

$$\Delta R=\frac{N}{V}sv_xF(v_x)\Delta v_x=\frac{N}{V}sv_x\sqrt{\frac{\alpha}{\pi}}\,e^{-\alpha v_x^2}\Delta v_x.$$

При каждом ударе стенка получает импульс $2m_0v_x$. Импульс, переданный участку стенки площадью s за единицу времени,

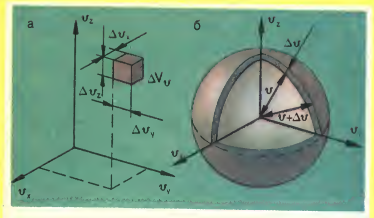


Рис. 4.8

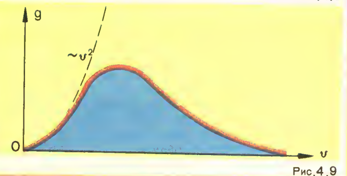


Рис. 4.9

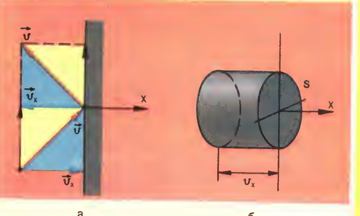


Рис. 4.10

есть сила, действующая на этот участок. Разделив силу на площадь участка s, найдем давление газа на стенку, обусловленное теми молекулами, у которых x-проекция скорости находится в интервале от v_x до $v_x+\Delta v_x$:

$$\Delta p = 2m_0 v_x \Delta R \frac{1}{s} = 2m_0 \frac{N}{V} \sqrt{\frac{\alpha}{\pi}}\, e^{-\alpha v_x^2} v_x^2 \Delta v_x. \quad (4.21)$$

Остается просуммировать, а точнее, проинтегрировать результат (4.21) по всем неотрицательным значениям скорости v_x:

$$p = 2m_0 \frac{N}{V} \sqrt{\frac{\alpha}{\pi}} \int_0^{\infty} e^{-\alpha v_x^2} v_x^2 dv_x. \quad (4.22)$$

Воспользуемся тем, что

$$\int_0^{\infty} e^{-\alpha v_x^2} v_x^2 dv_x = \frac{1}{4} \sqrt{\frac{\pi}{\alpha^3}}\,.$$

Таким образом,

$$p = m_0 N / 2\alpha V. \quad (4.23)$$

Распределение Максвелла проясняется окончательно. Мы долго испытывали терпение читателя, используя во всех предыдущих формулах таинственный параметр α. Из (4.23) видно, что $\alpha = m_0 N / 2pV$. Учитывая, что газ находится в состоянии теплового равновесия, воспользуемся уравнением Менделеева — Клапейрона: $pV = mRT/M$. Поскольку $R = N_A k$ (N_A — постоянная Авогадро, k — постоянная Больцмана, равна $1{,}38 \cdot 10^{-23}$ Дж/К) и, кроме того, $N_A m/M = N$, перепишем уравнение Менделеева — Клапейрона в виде:

$$pV = NkT. \quad (4.24)$$

Из (4.23) и (4.24) получаем:

$$\alpha = m_0 / 2kT. \quad (4.25)$$

Таким образом, выражение (4.19) принимает вид:

$$\Delta w_v = g(v)\Delta v = 4\pi \left(\frac{m_0}{2\pi kT}\right)^{3/2} e^{-\frac{m_0 v^2}{2kT}} v^2 \Delta v. \quad (4.26)$$

Температура как мера средней энергии молекул. Среднее значение квадрата скорости молекул идеального газа можно найти, используя интеграл (1.17) и применяя результат (4.26):

$$\langle v^2 \rangle = \int_0^{\infty} v^2 g(v) dv = 4\pi \left(\frac{m_0}{2\pi kT}\right)^{3/2} \int_0^{\infty} e^{-\frac{m_0 v^2}{2kT}} v^4 dv. \quad (4.27)$$

Учитывая, что

$$\int_0^\infty e^{-\alpha v^2} v^4\, dv = \frac{3}{8} \sqrt{\frac{\pi}{\alpha^5}},$$

получаем из (4.27):

$$\langle v^2 \rangle = \frac{3}{2\alpha} = \frac{3kT}{m_0}. \qquad (4.28)$$

Применяя модель идеального газа, можно пренебрегать энергией взаимодействия молекул друг с другом по сравнению с их кинетической энергией, т. е. можно представлять энергию молекулы как $\varepsilon = \frac{m_0 v^2}{2}$. Учитывая (4.28), получаем в итоге следующее выражение для средней энергии молекулы идеального газа:

$$\langle \varepsilon \rangle = \frac{m_0}{2} \langle v^2 \rangle = \frac{3}{2} kT. \qquad (4.29)$$

Мы видим, таким образом, что температура может рассматриваться в качестве *меры средней энергии молекул*.

Из (4.29) следует, что *внутренняя энергия* равновесного идеального газа, содержащего N молекул и имеющего температуру T, есть

$$U = \frac{3}{2} NkT. \qquad (4.30)$$

Молекулярно-кинетические представления позволили объяснить тот факт, что внутренняя энергия идеального газа пропорциональна его абсолютной температуре и не зависит от объема, занимаемого газом. Этот факт мы использовали при рассмотрении некоторых вопросов термодинамики.

Флуктуации

Флуктуации микровеличин и макровеличин. Под *микровеличинами* будем понимать величины, относящиеся к отдельной молекуле, а под *макровеличинами* — к макроскопическому объекту, например к газу в целом. Скорость v и энергия ε молекулы — это микровеличины; внутренняя энергия газа U, температура T, давление p — это макровеличины.

Будем мысленно следить за энергией какой-нибудь молекулы в газе. Энергия изменяется случайным образом от столкновения к столкновению. Зная функцию $\varepsilon(t)$ для достаточно большого промежутка времени τ, можно найти среднее значение энергии молекулы:

$$\langle \varepsilon \rangle = \frac{1}{\tau} \int_0^\tau \varepsilon(t)\, dt. \qquad (4.31)$$

Заметим, что в параграфе «Давление и температура идеального газа» мы подходили к понятию средней энергии иначе. Вместо

того чтобы следить за энергией какой-то молекулы в течение некоторого промежутка времени, мы фиксировали для какого-то момента времени энергии всех молекул и сумму делили на число молекул; именно таков смысл соотношения (4.27). Можно сказать, что здесь мы рассматривали *усреднение по коллективу*. Формула же (4.31) соответствует *усреднению по времени*. Оба усреднения приводят к одинаковым результатам.

Но вернемся к энергии молекулы в газе. С течением времени энергия $\varepsilon(t)$ случайным образом колеблется, или, как принято говорить, *флуктуирует* около среднего значения $\langle\varepsilon\rangle$. В качестве меры отклонения энергии от среднего значения может быть выбрана *дисперсия*

$$D(\varepsilon)=\langle\varepsilon^2\rangle-\langle\varepsilon\rangle^2. \tag{4.32}$$

Дисперсию $D(\varepsilon)$ называют *квадратичной флуктуацией* энергии ε. Зная распределение молекул по скоростям, можно вычислить $\langle\varepsilon^2\rangle$ по формуле:

$$\langle\varepsilon^2\rangle=\int_0^\infty\left(\frac{m_0v^2}{2}\right)^2 g(v)\,dv. \tag{4.33}$$

Подставляя сюда плотность вероятности $g(v)$ из (4.26), находим (математические выкладки для простоты опущены):

$$\langle\varepsilon^2\rangle=15(kT)^2/4. \tag{4.34}$$

С учетом (4.29) получаем:

$$D(\varepsilon)=\langle\varepsilon^2\rangle-\langle\varepsilon\rangle^2=\frac{3}{2}(kT)^2. \tag{4.35}$$

Отношение квадратного корня из квадратичной флуктуации к среднему значению величины называется *относительной флуктуацией* величины. В данном случае это отношение равно примерно единице:

$$\xi=\frac{\sqrt{D(\varepsilon)}}{\langle\varepsilon\rangle}=\sqrt{\frac{2}{3}}. \tag{4.36}$$

Размах флуктуаций микровеличины оказывается того же порядка, что и ее среднее значение.

Теперь рассмотрим флуктуации какой-нибудь макровеличины, например внутренней энергии газа, состоящего из N одноатомных молекул. Пусть $U(t)$ — значение внутренней энергии газа в момент t:

$$U(t)=\sum_{i=1}^{N}\varepsilon_i(t). \tag{4.37}$$

Эти значения флуктуируют около среднего значения $\langle U\rangle$. Флуктуации внутренней энергии газа могут быть связаны с хаотическими актами обмена энергией между молекулами газа и стен-

кой сосуда. Поскольку среднее от суммы есть сумма средних, то

$$\langle U\rangle=\sum_{i=1}^{N}\langle\varepsilon\rangle=N\langle\varepsilon\rangle. \qquad (4.38)$$

Мы воспользовались тем, что средняя энергия одинакова для любой молекулы.

Дисперсию $D(U)$ предварительно запишем в виде: $D(U)=\langle U^2\rangle-\langle U\rangle^2=\langle[U(t)-\langle U\rangle]^2\rangle$. Разность $U(t)-\langle U\rangle$ будем обозначать через δU, так что

$$D(U)=\langle(\delta U)^2\rangle. \qquad (4.39)$$

Используя (4.37) и (4.38), представим:

$$\delta U=U(t)-\langle U\rangle=\sum_{i=1}^{N}\varepsilon_i(t)-N\langle\varepsilon\rangle=\sum_{i=1}^{N}[\varepsilon_i(t)-\langle\varepsilon\rangle]=\sum_{i=1}^{N}\delta\varepsilon_i.$$

Таким образом,

$$D(U)=\left\langle\left(\sum_{i=1}^{N}\delta\varepsilon_i\right)^2\right\rangle. \qquad (4.40)$$

Надо возвести в квадрат сумму из N слагаемых, а затем усреднить каждое из слагаемых, которые получатся после возведения в квадрат. Возведение в квадрат дает N слагаемых вида $(\delta\varepsilon_i)^2 (i=1, 2, ..., N)$. Усредняя эти слагаемые, получим в итоге $N\langle(\delta\varepsilon)^2\rangle$. Кроме того, возведение в квадрат дает ряд слагаемых, называемых обычно перекрестными членами; это слагаемые вида $2\delta\varepsilon_i\delta\varepsilon_j$, где $i\neq j$. После усреднения каждое из этих слагаемых даст нуль. Действительно, $\langle\delta\varepsilon_i\delta\varepsilon_j\rangle=\langle\delta\varepsilon_i\rangle\langle\delta\varepsilon_j\rangle$. Что же касается средних $\langle\delta\varepsilon_i\rangle$ и $\langle\delta\varepsilon_j\rangle$, то они равны нулю, поскольку отклонения величины от ее среднего значения происходят равноправно в обе стороны. Итак,

$$D(U)=N\langle(\delta\varepsilon)^2\rangle=ND(\varepsilon). \qquad (4.41)$$

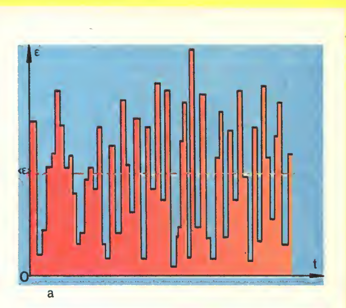
Используя (4.35), получаем в итоге следующее выражение для квадратичной флуктуации внутренней энергии газа:

$$D(U)=\frac{3}{2}N(kT)^2. \qquad (4.42)$$

Относительная флуктуация внутренней энергии есть

$$\xi=\frac{\sqrt{D(U)}}{\langle U\rangle}=\sqrt{\frac{2}{3}}\cdot\frac{1}{\sqrt{N}}. \qquad (4.43)$$

Мы видим, что относительная флуктуация внутренней энергии газа из N молекул пропорциональна $1/\sqrt{N}$, т. е. очень мала (напомним, что в кубическом сантиметре газа при нормальном давлении содержится около 10^{19} молекул). Для всех макровеличин $\xi\sim1/\sqrt{N}$, что и позволяет на практике пренебрегать их флуктуациями, рассматривая средние значения макровеличин в качестве



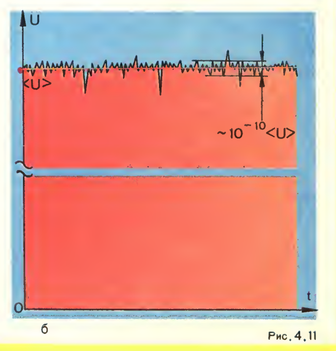


Рис. 4.11

истинных. На рисунке 4.11 сопоставляется характер флуктуаций для микровеличины ε и макровеличины U.

Итак, полная внутренняя энергия U не является фиксированной величиной для данного равновесного состояния макроскопического объекта. Она слегка изменяется во времени, испытывая небольшие флуктуации около своего среднего значения. Точно так же флуктуируют около своих средних значений температура, давление, энтропия.

Броуновское движение. Познакомившись с результатом (4.43), читатель, возможно, сделает вывод, что в обычных условиях, когда мы имеем дело с макроскопическими объектами и характеризующими их макровеличинами, флуктуации не проявляются. Однако наблюдать флуктуации можно фактически невооруженным глазом. В качестве примера отметим *броуновское движение.*

В 1827 году английский биолог Роберт Броун обнаружил при помощи обычного микроскопа хаотическое движение мелких частиц (пыльцы растений), взвешенных в воде. «Движение это, как я убежден,— писал Броун,— обусловлено не потоками в жидкости, не постепенным ее испарением, а принадлежит самим частицам». Правильное объяснение природы броуновского движения дал в 1905 году Альберт Эйнштейн (1879—1955). Он показал, что причиной броуновского движения является хаотическая бомбардировка взвешенных мелких частиц молекулами окружающей жидкости, находящимися в тепловом движении.

Представим себе взвешенный в жидкости маленький диск диаметром, например, около 10^{-4} см. Отнесенное к единице времени число ударов молекул жидкости по одной стороне диска равно в среднем числу ударов по другой стороне. Это равенство соблюдается только в среднем. В действительности же число ударов по одной стороне диска в течение какого-то малого промежутка времени может оказаться заметно больше числа ударов по другой стороне. В результате диск получит нескомпенсированный импульс и совершит в соответствующем направлении скачок. Можно сказать, что причиной движения диска являются *флуктуации давления,* оказываемого молекулами жидкости на разные стороны диска.

Эйнштейн рассмотрел конкретную физическую модель, выбрав в качестве броуновской частицы шарик. Он показал, что средний квадрат смещения такой частицы за время наблюдения τ описывается формулой:

$$\langle l^2 \rangle = \frac{\tau}{8\pi\eta r} kT, \quad (4.44)$$

где r — радиус шарика, η — коэффициент вязкости жидкости, T — ее температура.

Голубой цвет неба. Мы воспринимаем цвет неба вследствие рассеяния солнечных лучей в земной атмосфере. Мысленно выделим в воздушном пространстве атмосферы множество маленьких ячеек-кубиков с длиной ребра, соответствующей длине волны света (около $0{,}5 \cdot 10^{-4}$см). Хаотическое движение молекул воз-

духа приводит к тому, что число молекул в пределах кубика будет случайно изменяться от одного кубика к другому. Оно будет случайно изменяться и в пределах одного кубика, если производить наблюдения в разные моменты времени. На этих *флуктуациях плотности* воздуха, проявляющихся в достаточно малых объемах, и происходит рассеяние света.

Согласно современной теории, интенсивность света ΔI, рассеянного объемом воздуха ΔV на расстоянии R от наблюдателя, описывается соотношением:

$$\Delta I = a\frac{\Delta V}{R^2}\cdot\frac{1}{\lambda^4}kT, \qquad (4.45)$$

где λ — длина волны света, T — температура воздуха, a — некий множитель, который мы здесь не будем расшифровывать. Из (4.45) видно, что рассеяние света происходит тем интенсивнее, чем меньше длина волны ($\Delta I \sim 1/\lambda^4$). Поэтому спектр рассеянного в земной атмосфере света оказывается смещенным в *коротковолновую часть*, что и объясняет наблюдаемый *голубой* цвет неба.

Формула Найквиста. Из закона Ома следует, что если в электрической цепи нет электродвижущей силы, то ток в ней идти не будет. Это, однако, не совсем верно. Дело в том, что флуктуации, связанные с тепловым движением электронов в проводнике, приводят к возникновению флуктуационных токов, так что можно говорить о наличии флуктуационной электродвижущей силы. В 1927 году американский физик X. Найквист показал, что на концах проводника с сопротивлением R, имеющего температуру T, возникает *флуктуация напряжения* δV, средний квадрат которой описывается формулой:

$$\langle(\delta V)^2\rangle = 4RkT\ \Delta\nu, \qquad (4.46)$$

где $\Delta\nu$ — интервал частот, в пределах которого измеряются флуктуации напряжения.

Флуктуации электрических величин играют важную роль в современной аппаратуре. Они являются принципиально неустранимым источником шума в каналах связи и определяют предел чувствительности измерительных приборов. Наряду с флуктуациями, обусловленными тепловым движением электронов в проводниках, отметим еще один важный тип флуктуаций — флуктуации числа электронов, вылетающих из нагретого катода электронной лампы.

Флуктуации и температура. Обратим внимание читателя на выражения (4.35) и (4.42). Видно, что квадратичная флуктуация связана с абсолютной температурой: $\sqrt{D} \sim T$. Об этом же говорят и формулы (4.44) — (4.46). Связь между квадратичной флуктуацией физической величины и температурой имеет глубокий смысл. Чем выше абсолютная температура тела, тем сильнее флуктуируют его физические параметры.

Выше отмечалось, что температура тела может рассматриваться

как мера средней энергии частиц тела. Следует помнить, что это справедливо лишь при условии, что тело находится в тепловом равновесии. Если же состояние некоего коллектива частиц существенно неравновесно (предположим, что рассматривается космический ливень или пучок частиц из ускорителя), то в этом случае средняя энергия частиц уже не может измеряться температурой. Более общий подход к понятию температуры тела предполагает ее связь не со средней энергией частиц, а с флуктуациями физических параметров тела. При этом температура может рассматриваться как мера флуктуаций. Измеряя флуктуации, можно в принципе измерять абсолютную температуру тела. Для такой цели наиболее подходят флуктуации электрических величин.

Связь температуры с флуктуациями указывает, в частности, на то, что понятие температуры, строго говоря, не имеет аналога в механике Ньютона. Температура предполагает наличие вероятностных процессов, она выступает в качестве меры дисперсии случайных величин.

Энтропия и вероятность

От формулы работы газа при изотермическом расширении к формуле Больцмана. Предположим, что идеальный газ, имеющий массу m и температуру T, изотермически расширяется от объема V_1 до объема V_2. Согласно (4.6), работа, совершенная газом при таком расширении, есть $(mRT/M)\ln(V_2/V_1)$. При изотермическом расширении работа совершается за счет теплоты Q, отбираемой газом от окружающих тел. Следовательно,

$$Q=\frac{mRT}{M}\ln\frac{V_2}{V_1}. \qquad (4.47)$$

Используя для уравнения состояния идеального газа выражение (4.24), преобразуем (4.47) к виду:

$$Q=NkT\ln\frac{V_2}{V_1}, \qquad (4.48)$$

где N — число молекул в газе. С учетом (4.10) заключаем, что приращение энтропии газа есть

$$\Delta S= Nk\ln\frac{V_2}{V_1}. \qquad (4.49)$$

Изотермическое расширение газа — обратимый процесс. Возрастание энтропии в *обратимом* процессе не должно удивлять читателя: мы рассматриваем энтропию газа, а газ является здесь незамкнутой системой (он совершает работу над поршнем, получает теплоту извне). Такое же увеличение энтропии будет наблюдаться в *необратимом* процессе расширения газа от V_1 до V_2 в случае, когда газ представляет собой замкнутую систему. Этот необратимый процесс можно осуществить следующим образом. Предположим, что в теплоизолированном сосуде объема V_0 есть пере-

городка и весь газ находится сначала по одну сторону перегородки, занимая объем V_1. Перегородка убирается, и газ начинает расширяться в пустоту. Процесс расширения рассматриваем от момента, когда убрана перегородка, до момента, когда газ займет объем V_2. Приращение энтропии газа в данном процессе также описывается формулой (4.49).

Используя пример с расширением газа в пустоту, можно объяснить увеличение энтропии на основе *вероятностей*. Вероятность того, что молекула газа окажется в объеме V_1, равна, очевидно V_1/V_0. Вероятность того, что одновременно с первой в объеме V_1 окажется другая какая-то молекула, равна $(V_1/V_0)^2$. Вероятность же всем N молекулам собраться в объеме V_1 равна $(V_1/V_0)^N$. Обозначим через w_1 вероятность реализации состояния газа, когда все молекулы оказываются в объеме V_1, а через w_2 — когда все молекулы оказываются в объеме V_2. Первая вероятность равна $(V_1/V_0)^N$, а вторая $(V_2/V_0)^N$. Таким образом,

$$\frac{w_2}{w_1}=\left(\frac{V_2}{V_1}\right)^N \qquad (4.50)$$

Используя (4.50), получаем из (4.49):

$$\Delta S=Nk\ln\frac{V_2}{V_1}=k\ln\left(\frac{V_2}{V_1}\right)^N=k\ln\frac{w_2}{w_1}. \qquad (4.51)$$

Посредством довольно простых рассуждений мы пришли к важному результату — знаменитой формуле Больцмана.

Формула Больцмана. В 1872 году Людвиг Больцман опубликовал свою знаменитую формулу, согласно которой энтропия системы в некотором состоянии пропорциональна *логарифму вероятности состояния.* Коэффициент пропорциональности в этой формуле был уточнен позднее, его назвали *постоянной Больцмана.* В современной записи формула Больцмана выглядит так:

$$S=k\ln w. \qquad (4.52)$$

Результат (4.51) получается из (4.52), если положить $S_1=k\ln w_1$ и $S_2=k\ln w_2$ и принять $\Delta S=S_2-S_1$.

Предположим, что система состоит из двух подсистем, одна из которых находится в состоянии *1* — с энтропией S_1 и вероятностью w_1, а другая в состоянии *2* — с энтропией S_2 и вероятностью w_2. Обозначим энтропию и вероятность состояния системы через S и w. Энтропия обладает свойством аддитивности, поэтому

$$S=S_1+S_2. \qquad (4.53,\ a)$$

Рассматриваемое состояние реализуется, если одновременно первая подсистема оказывается в состоянии *1*, а вторая в состоянии *2*. Согласно теореме умножения вероятностей,

$$w=w_1w_2. \qquad (4.53,\ б)$$

Видно, что соотношения (4.53) согласуются с формулой Больцмана: $S=k\ln(w_1w_2)=k\ln w_1+k\ln w_2=S_1+S_2$.

Макросостояния и микросостояния системы. Уточним, что следует понимать под «вероятностью состояния системы» в формуле Больцмана. Для этого введем понятия *макросостояния* и *микросостояния.*

Рассмотрим простую систему, состоящую всего из четырех частиц, каждая из которых с равной вероятностью может находиться в одном из двух состояний. Можно представить себе сосуд, мысленно разделенный на две одинаковые половины (левую и правую), и всего четыре молекулы внутри него. Каждая из молекул с равной вероятностью может быть обнаружена в левой или правой половине. Возможны пять макросостояний данной системы: *1* — в левой половине нет ни одной молекулы, *2* — в левой половине одна молекула, *3* — в левой половине две молекулы, *4* — в левой половине три молекулы, *5* — в левой половине четыре молекулы. Различные макросостояния могут быть реализованы *разным числом равноправных способов,* иными словами, различным макросостояниям соответствуют разные числа микросостояний. Это видно на рисунке 4.12, где для разных молекул использованы различные цвета. Видно, что макросостояния *1* и *5* могут быть реализованы каждое одним способом. Каждому из них соответствует одно микросостояние. Макросостояниям *2* и *4* соответствуют по четыре микросостояния. Макросостоянию *3* соответствуют шесть микросостояний, это макросостояние может быть реализовано шестью равноправными способами. Всего в данном случае имеется 16 микросостояний. Все они равновероятны. Вероятность макросостояния *пропорциональна числу соответствующих ему микросостояний.* Именно эта вероятность и фигурирует в формуле Больцмана. Заметим, что число микросостояний, соответствующих данному макросостоянию, называют *статистическим весом* макросостояния.

Предположим, что в сосуде, разделенном на две половины, находятся не четыре, а N молекул. В данном случае имеется $N+1$ макросостояний, которые удобно обозначать числами 0, 1, 2, 3, ..., N — по числу молекул, находящихся, скажем, в левой половине. Статистический вес n-го макросостояния равен числу *сочетаний* из N по n:

$$C_N^n = \frac{N!}{(N-n)!\,n!}. \tag{4.54}$$

Это есть число микросостояний, соответствующих n-му макросостоянию. Полное число микросостояний описывается суммой $\sum\limits_{n=0}^{N} C_N^n$. Вероятность n-го макросостояния есть

$$w_n = C_N^n \Big/ \sum_{n=0}^{N} C_N^n. \tag{4.55}$$

Пример с использованием формулы Больцмана. Предположим, что газ, состоящий из N молекул, расширяется в пустоту. Его объем возрастает вдвое. Требуется найти увеличение энтропии газа.

$W_3 = 3/8$

$W_2 = 1/4$ $W_4 = 1/4$

$W_1 = 1/16$ $W_5 = 1/16$

1 2 3 4 5

Рис. 4.12

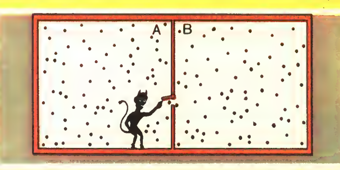


Рис. 4.13

Начальное состояние газа — макросостояние с $n=0$ (все молекулы в правой половине сосуда); *конечное* — макросостояние с $n=N/2$ (молекулы равномерно распределены по обеим половинам сосуда, что соответствует увеличению объема газа вдвое). Мы полагаем здесь, что N — четное число (при больших N эта оговорка несущественна). В соответствии с (4.54) и (4.55) представим:

$$\frac{w_{N/2}}{w_0}=\frac{C_N^{N/2}}{C_N^0}=C_N^{N/2}=\frac{N!}{(N/2)!\,(N/2)!}. \qquad (4.56)$$

Согласно формуле Больцмана, искомое увеличение энтропии газа есть

$$\Delta S=k\ln\frac{w_{N/2}}{w_0}=k\ln\frac{N!}{(N/2)!\,(N/2)!}. \qquad (4.57)$$

Учитывая, что N — очень большое число, воспользуемся приближенным соотношением:

$$\ln(N!)=N\ln N, \qquad (4.58)$$

после чего результат (4.57) принимает вид:

$$\Delta S=kN\ln 2. \qquad (4.59)$$

Этот же результат следует из (4.49), если принять $V_2/V_1=2$. **Энтропия как мера беспорядка в системе.** Вернемся к рисунку 4.12. Макросостояния *1* и *5* четко выявляют определенную структуру системы — ее разделение на две половины. В одной половине находятся молекулы, в другой молекул нет. Макросостояние *3*, напротив, совсем не выявляет указанной структуры, поскольку молекулы равномерно распределены по обеим половинам. Наличие определенной внутренней структуры связывают с *порядком* в системе, отсутствие структуры — с *беспорядком*. Чем выше степень упорядоченности макросостояния, тем меньше его статистический вес (тем меньше число соответствующих микросостояний). Разупорядоченные макросостояния, характеризующиеся отсутствием внутренней структуры, имеют большой статистический вес. Они могут быть реализованы многими способами, иначе говоря, многими микросостояниями.

Все это позволяет рассматривать энтропию как *меру беспорядка в системе*. Чем больше беспорядок в данном макросостоянии, тем больше его статистический вес и тем, следовательно, больше энтропия.

Статистическое объяснение второго начала термодинамики. Формула Больцмана позволяет очень просто объяснить постулируемое вторым началом термодинамики возрастание энтропии при необратимых процессах в замкнутой системе. *Возрастание энтропии* означает переход системы из *менее вероятных* в *более вероятные* состояния. Рассмотренный выше пример с расширением газа в пустоту иллюстрирует это утверждение: при расширении газ совершал переход из менее вероятных в более вероятные макросостояния.

Любые процессы в замкнутых системах протекают в таком направлении, чтобы энтропия системы при этом не убывала. Это означает, что всем реально протекающим процессам соответствуют переходы в более вероятные состояния или, в крайнем случае, переходы между равновероятными состояниями.

При вероятностном подходе энтропия выступает как мера беспорядка в системе. Закон возрастания энтропии в замкнутых системах есть, следовательно, закон *увеличения степени беспорядка* в этих системах. Иными словами, переход из менее вероятных в более вероятные состояния соответствует переходам «порядок→беспорядок». Когда, например, молот ударяет по наковальне, упорядоченная составляющая движения молекул молота, связанная с его поступательным движением как целого, переходит в неупорядоченное тепловое движение молекул наковальни и молота.

Количество энергии в замкнутой системе с течением времени не изменяется. Однако изменяется *качество* энергии. В частности, уменьшается ее способность совершать полезную работу. Возрастание энтропии в замкнутой системе есть, по сути дела, постепенное разрушение системы. Всякая замкнутая система со временем неизбежно разупорядочивается, разрушается. Изоляция системы отдает ее во власть разрушающих случайностей, которые всегда направляют систему по пути к беспорядку. Как выразился французский ученый Л. Бриллюэн, «второе начало термодинамики говорит о смерти вследствие изоляции».

Чтобы поддерживать или, тем более, повышать упорядоченность системы, надо ею *управлять*, для чего необходимо, прежде всего, чтобы *не было изоляции системы*, чтобы система не была замкнутой. Конечно, после того как система лишается «защитной оболочки», открывается доступ к ней для различного рода внешних дезорганизующих факторов. Однако наряду с этим открывается доступ и для управляющих факторов. Действие последних может приводить к уменьшению энтропии системы. Разумеется, это не противоречит второму началу термодинамики: понижение энтропии имеет локальный характер — уменьшается лишь энтропия данной системы. Это уменьшение с избытком компенсируется возрастанием энтропии в других системах, в частности тех, которые осуществляют управление данной системой.

Второе начало термодинамики и флуктуации. Вероятностный подход не только объяснил второе начало термодинамики, но и показал, что требования этого закона не являются категорическими. Диктуемое вторым началом направление развития процессов не является жестко предопределенным. Это есть лишь *наиболее вероятное* направление. Нарушения второго начала термодинамики в принципе допустимы. Обычно мы не наблюдаем их лишь потому, что они *маловероятны*.

Газ самопроизвольно расширяется в пустоту. Это наиболее вероятное направление процесса. Однако в принципе возможна ситуация, когда скорости молекул в газе окажутся вдруг ориентированными таким образом, чтобы газ самопроизвольно сжался. Такая ситуа-

ция маловероятна. Ее исключительно малая вероятность связана с огромным числом молекул в любом макрообъеме газа. Самопроизвольное сжатие газа следует рассматривать как флуктуацию его плотности. Чем больше молекул в газе, тем, как известно, меньше характерная величина относительной флуктуации (напомним, что она пропорциональна $1/\sqrt{N}$), тем, следовательно, маловероятнее наблюдать такую флуктуацию в масштабах макромира. Допустим, что рассматриваемое явление требует участия относительно небольшого числа молекул. В этом случае уже нетрудно наблюдать различного рода флуктуации, свидетельствующие о нарушениях второго начала термодинамики. В предыдущем параграфе мы говорили о флуктуациях плотности воздуха в пределах достаточно малого объема, линейные размеры которого соответствовали длине волны света. Эти флуктуации проявляются в самопроизвольных сжатиях и разрежениях воздуха, которые и обусловливают наблюдаемый нами голубой цвет неба.

Для броуновской частицы наиболее вероятно получить в единицу времени одинаковое число ударов молекул жидкости с той и с другой стороны. Однако, вследствие малости размеров броуновской частицы, вполне вероятны флуктуации давления, обусловленные нескомпенсированностью ударов с разных сторон, такие, чтобы частица совершила скачок в некотором направлении. Совершая очередной скачок, броуновская частица наглядно демонстрирует самопроизвольное превращение теплоты, отбираемой от жидкости, в кинетическую энергию своего поступательного движения.

Мы видим, таким образом, что вероятностная трактовка энтропии, а вместе с тем и второго начала термодинамики отвечает более глубокому пониманию природы процессов в макросистемах. Вероятностный подход не только объясняет все те загадки, на которые термодинамика не могла дать ответа. Этот подход дает больше — он показывает, что второе начало термодинамики само имеет *вероятностную природу,* что оно выполняется лишь в среднем, что всевозможные флуктуации непрерывно нарушают этот закон термодинамики. Мы приходим к важному выводу: *в основе второго начала термодинамики лежат не жестко детерминированные, а вероятностные закономерности.*

Энтропия и информация

Связь между энтропией и информацией. В третьей главе было показано, что понятие информации основано на вероятности. Теперь мы убедились, что вероятность лежит и в основе энтропии. Единство природы *информации* и *энтропии* оказывается не случайным. Увеличение энтропии соответствует переходу системы из менее упорядоченных в более упорядоченные состояния. Такой переход сопровождается уменьшением информации, содержащейся в структуре системы. Беспорядок, неопределенность можно рассматривать, как недостачу информации. В свою очередь, информация есть не что иное, как уменьшение неопределенности.

Согласно второму началу термодинамики, энтропия замкнутой системы увеличивается с течением времени. Этот процесс соответствует рассматривавшемуся в третьей главе процессу потери информации в результате действия случайных факторов. Флуктуации физических параметров обусловливают случайные нарушения второго начала термодинамики. Наблюдаются случайные понижения энтропии. Эти процессы соответствуют обсуждавшимся ранее явлениям генерации информации из шума. Воздействуя определенным образом на систему, можно понизить ее энтропию (за счет повышения энтропии других систем). Это есть процесс управления, который требует использования определенной информации.

Все это говорит о существовании связи между информацией и энтропией. Впервые на эту связь указал в 1929 году венгерский физик Л. Сцилард.

Итак, если *энтропия — мера беспорядка*, неопределенности в системе, то *информация* — напротив, *мера порядка*, структурной определенности. Возрастанию информации соответствует уменьшение энтропии, и наоборот, уменьшению информации отвечает увеличение энтропии.

Формула Больцмана и формула Хартли. Ранее мы познакомились с формулой Хартли (см. 3.1). Согласно этой формуле, информация, требуемая для выявления одного из N_1 равновероятных исходов, есть $I=\log_2 N_1$. Пусть N_1 — число путей на железнодорожной станции. Диспетчеру надо послать сигнал, указывающий тот путь, на который следует принять приближающийся к станции поезд. Подавая сигнал, диспетчер производит выбор среди N_1 равновероятных исходов. Этот сигнал содержит информацию $I_1=\log_2 N_1$. Допустим далее, что некоторые пути ремонтируются, так что диспетчер должен выбирать уже среди N_2 исходов ($N_2<N_1$). В этом случае его сигнал содержит информацию $I_2=\log_2 N_2$. Разность

$$\Delta I=I_1-I_2=\log_2\frac{N_1}{N_2} \qquad (4.60)$$

есть информация о ремонте определенных путей. Иными словами, это есть информация, требуемая для уменьшения числа равновероятных исходов от N_1 до N_2.

Сопоставим существование N равновероятных исходов с наличием N равновероятных микросостояний, т. е. со статистическим весом N некоторого макросостояния. Согласно формуле Больцмана, уменьшение статистического веса макросостояний от N_1 до N_2 означает, что энтропия системы получила приращение

$$\Delta S=-k\ln\frac{N_1}{N_2}. \qquad (4.61)$$

Мы используем здесь знак минус, поскольку при уменьшении статистического веса энтропия уменьшается (приращение является отрицательным). В соответствии с (4.60) для реализации рассматриваемого отрицательного приращения энтропии требуется

приращение информации $\Delta I = \log_2(N_1/N_2)$. Сопоставляя (4.60) и (4.61) и учитывая при этом, что $\log_2(N_1/N_2) = \frac{\ln(N_1/N_2)}{\ln 2}$, получаем:

$$\Delta S = -\Delta I \frac{k}{\ln 2}. \qquad (4.62)$$

Таким образом, приращению информации ΔI соответствует уменьшение энтропии системы, равное $\Delta I k / \ln 2$.

Согласно высказыванию Н. Винера, «информация — это отрицательная энтропия». Л. Бриллюэн предложил использовать вместо термина «отрицательная энтропия» термин «негэнтропия».

«Демон» Максвелла и его изгнание. В 1871 году Максвелл сформулировал следующую умозрительную ситуацию, выглядевшую как парадокс. Пусть сосуд с газом разделен на две половины (A и B) перегородкой, в которой имеется небольшое отверстие с клапаном. Предположим, рассуждал Максвелл, что некое «существо» (он назвал его «демоном») управляет клапаном, закрывая и открывая отверстие таким образом, чтобы пропустить наиболее быстрые молекулы из половины A сосуда в половину B, а наиболее медленные молекулы из половины B в половину A. В результате «демон» повысит температуру в половине B и понизит ее в половине A, не производя при этом работы, что, очевидно, противоречит второму началу термодинамики.

Глядя на условный рисунок с изображением «демона» Максвелла (рис. 4.13), читатель, разумеется, не должен думать о нечистой силе. Речь идет о некотором устройстве открывания и закрывания отверстия, которое могло бы действовать подобно описанному выше «демону».

В принципе можно было бы предложить три типа таких устройств. Первый тип — устройство, управляемое молекулами газа, находящегося в сосуде. Можно вообразить, что имеется дверца, открывающаяся в одну сторону и реагирующая на энергию ударяющихся о нее молекул; быстрые молекулы открывают дверцу, а медленные не открывают. Чтобы открываться от удара отдельных молекул, дверца должна быть фантастически легкой. Однако подобная дверца, если бы даже удалось ее изготовить, не смогла бы выполнять функции «демона». На такую дверцу действовали бы в равной мере как флуктуации, связанные с движением молекул газа, так и флуктуации, связанные с тепловым движением молекул вещества, из которого сделана дверца. Эта дверца действовала бы хаотическим образом и не могла бы осуществить сортировку молекул.

Второй тип «демона» — устройство, управляемое извне. Допустим, что каким-то образом удалось осуществить контроль над молекулами, подлетающими к отверстию в перегородке. Контрольное устройство подает в нужный момент тот или иной сигнал, и клапан открывается или, напротив, закрывается. Если не рассматривать технические проблемы, то следует признать, что такой спо-

соб сортировки молекул в принципе возможен. Однако он не годится, поскольку «демон» Максвелла должен действовать в *замкнутой* системе. Это принципиально, так как именно в замкнутой системе понижение энтропии противоречит второму началу. У нас же система незамкнутая. Наш «демон» получает информацию извне. Получение информации может рассматриваться как приток в систему отрицательной энтропии (негэнтропии), что эквивалентно уменьшению энтропии системы.

Остается еще один тип «демона» — «демон» в виде некоего *разумного* существа. Однако и этот вариант не годится, поскольку, как выразился Эйнштейн, «в среде, находящейся в равновесии, разумный механизм не может действовать». Иными словами, в замкнутой равновесной системе жизнь, а тем более разум невозможны.

Энтропия и жизнь. Живой организм — в высшей степени упорядоченная система с низкой энтропией. Существование живого организма предполагает непрерывное поддержание энтропии системы на низком уровне, непрерывное противодействие разупорядочивающим факторам и, в частности, факторам, вызывающим заболевания. Может показаться, что живой организм не подчиняется требованиям второго начала.

Это, конечно, не так. Необходимо учитывать, что любой живой организм — это *незамкнутая* система, пребывающая в существенно *неравновесном* состоянии. Эта система активным образом взаимодействует с окружающей средой, непрерывно черпая из нее негэнтропию. Известно, например, что пища имеет более низкую энтропию, нежели отходы.

Человек не просто живет. Он трудится, творит и, следовательно, активно понижает энтропию. Все это возможно лишь благодаря тому, что человек получает необходимое количество негэнтропии (информации) из окружающей среды. Она поступает к нему по двум различным каналам. Первый связан с процессом обучения. Второй связан с физиологическими процессами обмена, происходящими в системе «человек + окружающая среда».

ГЛАВА 5

Вероятность в микромире

Сегодня квантовая теория привела нас к более глубокому пониманию: она установила более тесную связь между статистикой и основами физики. Это является событием в истории человеческого мышления, значение которого выходит за пределы самой науки.

М. Борн

...Квантовая механика позволила утверждать о существовании в законах природы первичных вероятностей.

В. Паули

Спонтанные микропроцессы

Классическая физика исходила из того, что случайное проявляется лишь в больших коллективах, например в коллективе молекул в макрообъеме газа. В поведении же отдельной молекулы классическая физика не усматривала элементов случайного. Исследования, в результате которых возникла и сформировалась *квантовая физика,* показали, что такая точка зрения не соответствует действительности. Оказалось, что случайность обнаруживает себя не только в коллективе, но и в поведении отдельного микрообъекта. Это хорошо демонстрируют *спонтанные* (*самопроизвольные*) микропроцессы.

Распад нейтрона. Характерный пример спонтанного микропроцесса — распад свободного нейтрона. Обычно нейтроны находятся в связанном состоянии. Вместе с протонами они выполняют роль «кирпичиков», из которых построены атомные ядра. Однако нейтроны можно наблюдать и вне ядер — в свободном состоянии. Свободные нейтроны образуются, например, в результате деления ядер урана.

Оказывается, что свободный нейтрон может *случайно, без какого-либо воздействия на него* превращаться в три частицы — протон, электрон и антинейтрино (точнее, электронное антинейтрино). Это превращение принято называть *распадом* нейтрона и записывать так:

$$n \to p + e^- + \bar{\nu}_e,$$

где n — нейтрон, p — протон, e^- — электрон, $\bar{\nu}_e$ — антинейтрино. Заметим, что термин «распад» не вполне удачен, так как наводит на мысль, будто нейтрон состоит из протона, электрона, антинейтрино. В действительности же все эти три частицы рождаются в момент гибели нейтрона, бессмысленно было бы искать их «внутри нейтрона».

В самом факте самопроизвольного распада нейтрона налицо

случайность. Вместе с тем здесь диалектически проявляется и *необходимость*. Чтобы ее обнаружить, надо рассмотреть большое число нейтронов. Пусть в момент $t=0$ имеется N_0 нейтронов в некотором объеме, причем $N_0 \gg 1$. Будем измерять число нейтронов в объеме в разные моменты времени t, в результате мы сможем построить функцию $N(t)$, график которой имеет вполне определенный вид (рис. 5.1). Это есть функция

$$N(t) = N_0 e^{-at}, \tag{5.1}$$

где a — некоторая постоянная. Ее обычно записывают как $1/\tau$. Измерения показывают, что $\tau = 10^3$ с.

Величину τ называют *временем жизни* нейтрона. Это название условно. Величина τ не является действительным временем жизни нейтрона. Это есть время, в течение которого число уцелевших (нераспавшихся) нейтронов уменьшается в e раз. Действительно, согласно (5.1) $N(\tau)/N_0 = e^{-\tau/\tau} = 1/e$. Что же касается действительного времени жизни нейтрона, то оно может в каждом конкретном случае существенно отличаться от τ как в ту, так и в другую сторону. Принципиально нельзя предсказать, когда распадется тот или иной нейтрон. Можно лишь говорить о *вероятности* нейтрону прожить до распада то или иное время. При достаточно большом числе нейтронов отношение $N(t)/N_0$ есть вероятность нейтрону не распасться в течение промежутка времени t. Из (5.1) следует, что эта вероятность равна $e^{-t/\tau}$.

Здесь следует обратить внимание на одну любопытную деталь. Когда мы говорим о вероятности нейтрону уцелеть в течение промежутка времени t, мы отнюдь не предполагаем, что этот промежуток измеряется от момента рождения нейтрона. Несущественно, сколько времени тот или иной нейтрон уже прожил к моменту $t=0$. Все равно вероятность уцелеть к моменту t будет равна $e^{-t/\tau}$. Можно сказать, что нейтроны «не стареют». Это означает, в частности, что было бы бессмысленно искать причины распада конкретного нейтрона внутри самого нейтрона, в каком-то «внутреннем механизме».

Любопытно, что выражающий необходимость закон (5.1) есть не что иное, как прямое следствие того, что акты распада происходят независимым образом, случайно. Вследствие случайности уменьшение числа нейтронов (иначе говоря, число распадов) ΔN за промежуток времени от t до $t+\Delta t$ пропорционально только числу нейтронов $N(t)$ в рассматриваемый момент и длительности Δt промежутка времени: $\Delta N = -aN(t)\Delta t$. Перепишем это равенство в виде $\frac{\Delta N}{\Delta t} = -aN(t)$. Переходя к пределу при $\Delta t \to 0$, получаем дифференциальное уравнение, известное как *уравнение экспоненциального убывания:*

$$\frac{dN}{dt} = -aN(t). \tag{5.2}$$

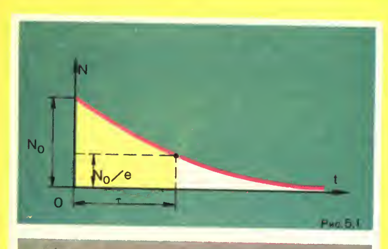

N
N_0
N_0/e
0
τ
t

Рис. 5.1

A
Z
237
233
229
225
221
217
213
Np^{237}
$2{,}2 \cdot 10^6$ лет
Pa^{233}
U^{233}
$1{,}6 \cdot 10^5$ лет
Th^{229}
27 суток
7340 лет
Ra^{225}
10 суток
Ac^{225}
15 суток
Fr^{221}
4,8 мин
At^{217}
47 мин
0,018 с
Bi^{213}
Po^{213}
$4{,}2 \cdot 10^{-6}$ с
Tl^{209}
Pb^{209}
Bi^{209}
3,3 час
2,2 мин
ТАЛЛИЙ
СВИНЕЦ
ВИСМУТ
ПОЛОНИЙ
АСТАТ
ФРАНЦИЙ
РАДИЙ
АКТИНИЙ
ТОРИЙ
ПРОТАКТИНИЙ
УРАН
НЕПТУНИЙ
81
82
83
84
85
86
87
88
89
90
91
92
93

Рис. 5.2

Функция (5.1) есть решение этого уравнения, удовлетворяющее начальному условию $N(0) = N_0$.

В заключение заметим, что если нейтрон не свободен, а связан мощными ядерными силами вместе с другими нейтронами и протонами в атомном ядре, то он утрачивает способность к распаду. Впрочем, в отдельных случаях он эту способность сохраняет. И тогда наблюдается явление бета-радиоактивности, о котором мы поговорим позже.

Нестабильность элементарных частиц. Нейтрон — отнюдь не единственная элементарная частица, способная спонтанно превращаться в другие частицы. Этим свойством, которое можно назвать *нестабильностью*, обладает подавляющее большинство элементарных частиц. Имеется лишь несколько стабильных частиц — фотон, нейтрино, электрон, протон.

Исследуя нестабильность различных частиц, можно обнаружить дополнительные проявления фактора случайного. Для примера возьмем частицу, называемую сигма-плюс-гипероном Σ^+. Она имеет положительный электрический заряд, равный по модулю заряду электрона, и массу в 2328 раз больше массы электрона. Как и нейтрон, эта частица самопроизвольно распадается. Ее время жизни (понимаемое так же, как и для нейтрона) равно $0{,}8 \cdot 10^{-10}$ с. В отличие от нейтрона, у гиперона есть не один, а два возможных способа распада:

либо $\Sigma^+ \to p + \pi^0$, либо $\Sigma^+ \to n + \pi^+$

(π^0 и π^+ — соответственно нейтральный и положительно заряженный пионы). Примерно в половине случаев гиперон распадается по одной схеме, а в половине случаев — по другой. Нельзя однозначно предсказать не только момент распада гиперона, но и схему распада.

Нестабильность атомных ядер (радиоактивность). Каждому химическому элементу соответствует не один, а несколько типов атомных ядер. Они содержат по одинаковому числу протонов (равному порядковому номеру данного химического элемента в периодической системе), но разному числу нейтронов; эти ядра называют *изотопами*. Большинство изотопов данного элемента *нестабильны*, или, иначе, *неустойчивы*. Неустойчивые изотопы данного элемента самопроизвольно превращаются в изотопы другого элемента с одновременным испусканием некоторых частиц. Это явление называют *радиоактивностью*. Радиоактивность была впервые обнаружена французским физиком Антуаном Анри Беккерелем в 1896 году. Термин «радиоактивность» был введен французским физиком Пьером Кюри, который совместно со своей женой Марией Склодовской исследовал это явление.

Исследования показали, что время жизни неустойчивых изотопов существенно различно для разных изотопов и разных способов превращений (разных типов радиоактивности). Оно может составлять миллисекунды, но может измеряться годами и даже столетиями. Встречаются изотопы с временем жизни свыше 10^8 лет.

Изучение долгоживущих нестабильных изотопов в природе позволяет ученым определять возраст пород.

Отметим различные типы радиоактивности. При этом будем обозначать через Z число протонов в ядре (порядковый номер химического элемента), а через A сумму числа протонов и нейтронов в ядре (так называемое массовое число). Одним из типов радиоактивности является *альфа-радиоактивность*, или, иначе, *альфа-распад*. При этом исходное ядро $(Z;\ A)$ распадается на альфа-частицу (ядро атома гелия, которое состоит из двух протонов и двух нейтронов) и ядро с числом протонов, равным $Z-2$, и массовым числом $A-4$:

$$X(Z;\ A) \to \alpha(2;\ 4) + Y(Z-2;\ A-4).$$

Другой тип радиоактивности — *бета-радиоактивность* (*бета-распад*). В этом случае один из нейтронов исходного атомного ядра превращается в протон, электрон и антинейтрино, подобно тому как это происходит со свободным нейтроном. Протон остается внутри нового ядра, а электрон и антинейтрино вылетают. Схема бета-распада может быть записана в виде:

$$X\ (Z;\ A) \to Y(Z+1;\ A) + e^- + \bar{\nu}_e.$$

Наблюдается также *протонная радиоактивность:*

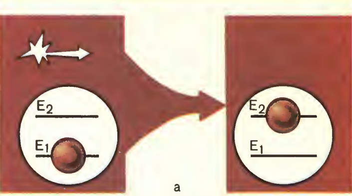
$$X(Z;\ A) \to p + Y\ (Z-1;\ A-1).$$

Кроме того, отметим *спонтанное деление* атомного ядра. Исходное ядро самопроизвольно распадается на два примерно одинаковых по массе «осколка» (два новых ядра), при этом рождаются несколько свободных нейтронов.

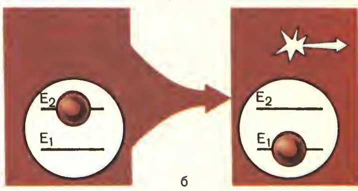
В качестве примера на рисунке 5.2 показана цепочка последовательных самопроизвольных превращений, в результате которых изотоп нептуния $^{237}\mathrm{Np}(Z=93;\ A=237)$ превращается в конечном счете в стабильный изотоп висмута $^{209}\mathrm{Bi}(Z=83;\ A=209)$. Цепочка состоит из «звеньев», отвечающих альфа-распадам (синие стрелки на рисунке) или бета-распадам (красные стрелки). Около стрелок указаны соответствующие значения времени жизни, понимаемые в вероятностном смысле. Такие цепочки называют *радиоактивными семействами.*

Вынужденные и спонтанные переходы в атоме. Читатель знает, что энергия атома принимает набор определенных (для данного атома) дискретных значений, в связи с чем говорят о наборе *энергетических уровней* атома. Если мы возбуждаем атомы, облучая их светом, то они переходят скачкáми с нижних энергетических уровней на верхние. Возбужденные атомы возвращаются на нижние уровни, испуская при этом свет. Такие скáчки атомов с одних энергетических уровней на другие называют *квантовыми переходами.*

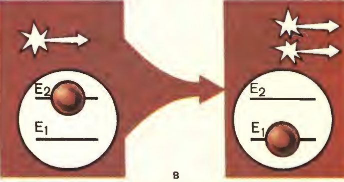
Если квантовый переход совершается в результате воздействия на атом, то его называют *вынужденным* (*индуцированным*). Если же переход совершается самопроизвольно, его называют *спон-*



а
Поглощение света



б
Спонтанное испускание света



в
Вынужденное испускание света

Рис. 5.3

танным. Переходы, связанные с возбуждением атома,— всегда вынужденные. Обратные переходы могут быть как вынужденными, так и спонтанными.

Рассмотрим для простоты всего два уровня энергии атома — с энергиями E_1 и E_2 (рис. 5.3). Переход $E_1 \to E_2$ есть вынужденный переход. Он происходит в результате поглощения атомом фотона, имеющего энергию $\varepsilon_{12} = E_2 - E_1$. Атом может возвратиться на уровень E_1 либо самопроизвольно, либо вынужденным образом. При этом испускается фотон с энергией ε_{12}. Спонтанный переход $E_2 \to E_1$ есть случайное событие. Вынужденный переход $E_2 \to E_1$ вызывается пролетающим вблизи атома фотоном. Энергия этого фотона должна быть равна ε_{12}. На рисунке показаны все три процесса: а) поглощение атомом фотона с энергией ε_{12} (при этом атом совершает переход $E_1 \to E_2$); б) спонтанное испускание атомом фотона с энергией ε_{12} (атом совершает переход $E_2 \to E_1$); в) вынужденное испускание атомом фотона с энергией ε_{12} при взаимодействии атома с первичным фотоном, также имеющим энергию ε_{12} (атом совершает переход $E_2 \to E_1$).

Надо отметить, что вынужденно испущенный фотон как бы копирует все свойства первичного фотона, вызвавшего переход атома. В частности, он будет иметь такое же направление движения, что и первичный фотон.

Как возникает генерация излучения в лазере? Во многих популярных книгах, посвященных лазерам, разъясняется роль вынужденного испускания фотонов, связанного с одновременным высвечиванием большого числа специально подобранных атомов или молекул (их называют *активными центрами*). Вынужденно испущенные фотоны движутся в едином направлении, образуя генерируемое лазером излучение.

Объяснение того, как возникает генерация в лазере, дается обычно по следующей схеме. Сначала каким-то способом (например, в результате облучения светом специальной мощной лампы) возбуждают активные центры. Добиваются, чтобы их число на верхнем энергетическом уровне оказалось больше, чем на нижнем. Тогда при появлении фотонов с энергией, равной разности энергий верхнего и нижнего уровней активного центра, процессы вынужденного испускания фотонов активными центрами будут происходить чаще, нежели обратные процессы (процессы поглощения фотонов). Это нетрудно понять, если учесть, что каждый первичный фотон с равной вероятностью может вызвать переход активного центра как снизу вверх (процесс поглощения света), так и сверху вниз (вынужденное испускание). Поэтому все определяется тем, где больше активных центров — вверху или внизу. Если их больше на верхнем уровне, то чаще будут происходить переходы сверху вниз, т. е. будут преобладать процессы вынужденного испускания. В результате и возникает мощный поток единым образом движущихся фотонов, представляющий собой излучение лазера.

В этом объяснении все верно. Однако при этом не рассматривается вопрос, откуда появляются те самые первичные фотоны, которые,

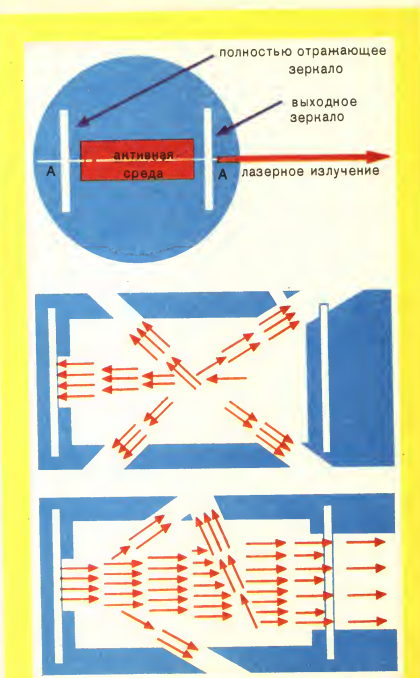

полностью отражающее
зеркало
выходное
зеркало
активная
среда
А
А
лазерное излучение

Рис. 5. 4

вызывая индуцированное испускание новых фотонов, как бы развязывают процесс лазерной генерации. А появляются эти фотоны за счет *спонтанных* переходов активных центров с верхнего уровня на нижний. Отмечая принципиально важную для работы лазера роль вынужденного испускания фотонов, надо в то же время не забывать о *первичности* (и в этом смысле фундаментальности) процессов спонтанного испускания. На этом можно было бы и закончить разговор о лазере. Но читатель хочет задать автору несколько вопросов.

Читатель. Вы говорили, что вынужденно испущенный фотон копирует все свойства, и в частности направление движения, первичного фотона.

Автор. Совершенно верно.

Читатель. Но ведь в спонтанных переходах рождаются фотоны со случайными направлениями движения. Следовательно, вынужденно испущенные фотоны также должны иметь разные направления движения. Один спонтанно родившийся фотон, пролетая вблизи множества возбужденных активных центров, вызовет появление лавины вынужденных фотонов в направлении своего движения. Второй спонтанно родившийся фотон поведет лавину вынужденных фотонов в другом направлении. Третий — в третьем направлении и так далее. Как же возникает лазерный луч?

Автор. Вы поставили очень важный вопрос. Обозначим через *AA* направление луча (рис. 5.4). Активная среда лазера имеет форму, вытянутую вдоль прямой *AA*. На этой прямой помещают два зеркала, одно из которых частично пропускает излучение. Фотоны, случайно родившиеся в направлении *AA* (или достаточно близком к нему), будут проходить в активной среде относительно большой путь, увеличенный за счет многократных отражений от зеркал. Взаимодействуя с возбужденными активными центрами, эти фотоны обусловят в конечном счете возникновение мощного потока вынужденно испущенных фотонов, образующего лазерный луч. Что же касается фотонов, которые случайно родились в иных направлениях, то они (и соответствующие вынужденно испущенные фотоны) пройдут в активной среде относительно короткий путь и быстро «выйдут из игры». Это хорошо видно на рисунке.

Отметим, что задающие направление лазерного луча зеркала образуют *резонатор* лазера.

Читатель. Получается, что лазерное излучение возникает из шума (из спонтанного излучения) благодаря *избирательности* усиления, т. е. благодаря тому, что усиление осуществляется главным образом вдоль определенного направления.

Автор. Именно так. Здесь мы встречаемся с уже знакомой ситуацией *отбора информации из шума*. Упорядоченное (когерентное) лазерное излучение как бы отбирается из шума. Этот отбор осуществляют зеркала резонатора. *Усиление* отбора происходит за счет вынужденного испускания — когда вторичный фотон копирует свойства первичного.

От соотношений неопределенностей к волновой функции

На примере спонтанных микропроцессов мы убедились, что в микромире случайное обнаруживает себя уже в поведении отдельного объекта. Это подводит нас вплотную к разговору о *первичности* и *фундаментальности* понятия вероятности в квантовой механике. Мы начнем его с *соотношений неопределенностей*, предложенных в 1927 году видным немецким физиком В. Гейзенбергом.

Соотношения неопределенностей. Микрообъект, движущийся по законам квантовой механики, не имеет, строго говоря, траектории движения. Это связано с тем, что микрообъект не может иметь *одновременно* и определенный импульс, и определенные координаты. Допустим, что микрообъект имеет определенную x-проекцию импульса. Оказывается, что в этом состоянии микрообъекта его x-координата не имеет какого-либо определенного значения. Другой крайний случай соответствует состоянию микрообъекта, в котором, напротив, его x-координата имеет определенное значение, а x-проекция импульса не имеет какого-либо определенного значения. Между отмеченными крайними случаями находится бесчисленное множество промежуточных, когда и x-координата, и x-проекция импульса не являются определенными, но их значения оказываются при этом в пределах каких-то интервалов. Пусть Δx — интервал, в пределах которого находятся значения x-координаты; будем называть Δx *неопределенностью* x-координаты. Соответственно будем рассматривать неопределенность x-проекции импульса Δp_x. Как впервые показал Гейзенберг, неопределенности Δx и Δp_x связаны соотношением:

$$\Delta x \Delta p_x \approx \hbar, \qquad (5.3)$$

где $\hbar = 1{,}05 \cdot 10^{-34}$ Дж·с — постоянная Планка. Аналогичные соотношения могут быть записаны и для других составляющих координаты и импульса микрообъекта: $\Delta y \Delta p_y \approx \hbar$, $\Delta z \Delta p_z \approx \hbar$.

Перед нами знаменитые *соотношения неопределенностей* Гейзенберга. Мы ограничимся здесь рассмотрением соотношений неопределенностей лишь для *координаты — импульса*. Заметим, однако, что подобные соотношения существуют также для некоторых других пар величин, например для *энергии — времени, угла — момента импульса*. Как писал Гейзенберг, «мы не можем интерпретировать процессы в атомарной области так же, как процессы большого масштаба. Если же мы пользуемся привычными понятиями, то их применимость ограничивается соотношениями неопределенностей».

Обсуждая соотношения неопределенностей, будем далее обращаться лишь к соотношению (5.3). Не следует думать, будто это соотношение указывает на невозможность осуществить сколь угодно точное измерение импульса или координаты микрообъекта. Оно утверждает лишь, что микрообъект не может иметь одновременно сколь угодно точно определенную координату и сколь угодно точно

определенный импульс. Стремясь, например, точнее фиксировать x-координату микрообъекта (иначе говоря, стремясь уменьшить Δx), мы неизбежно будем переводить микрообъект в такие состояния, в которых x-проекция его импульса будет характеризоваться все большей неопределенностью. В пределе, когда x-координата микрообъекта имеет определенное значение (микрообъект точно локализован), неопределенность x-проекции его импульса становится сколь угодно большой. Наоборот, стремясь точнее фиксировать x-проекцию импульса микрообъекта, мы неизбежно будем переводить его в состояния, в которых x-координата будет все более неопределенной.

Рассмотрим плоскость, где вдоль одной оси откладываются значения x-координаты объекта (ось x), а вдоль другой — значения x-проекции импульса (ось p_x) (рис. 5.5). Если бы объект подчинялся законам классической механики, то любое его состояние представлялось бы на этой плоскости в виде некоторой точки. Состоянию же микрообъекта соответствует на этой плоскости некоторый прямоугольник площадью $\hbar$. Возможны разные типы состояний микрообъекта. Им соответствуют прямоугольники разной формы. Некоторые из них изображены на рисунке.

Соотношения неопределенностей и волновые свойства микрообъекта. В 1924 году знаменитый французский физик Луи де Бройль выдвинул гипотезу, согласно которой микрообъект обладает свойствами не только *корпускулы*, но и *волны*. Его корпускулярные характеристики (энергия ε, импульс p) связаны, по де Бройлю, с волновыми характеристиками (частотой ω, длиной волны λ) соотношениями:

$$\varepsilon = \hbar\omega, \; p = 2\pi\hbar/\lambda. \tag{5.4}$$

Эта гипотеза представлялась многим физикам абсурдной. Было совершенно непонятно, что именно скрывается под понятием «длина волны частицы».

Но вот в 1927 году в опытах по прохождению пучка электронов сквозь тонкие металлические пластинки был получен потрясающий результат: рассеянные пластинкой электроны демонстрировали картину *дифракционных колец* (рис. 5.6). *Дифракция электронов* на кристаллической решетке оказалась экспериментально подтвержденным фактом! Явления дифракции и интерференции всегда связывали с наличием каких-то волн. Поэтому опыты по дифракции электронов были единодушно восприняты, как доказательство существования у электрона волновых свойств. Природа «электронных волн» оставалась по-прежнему загадочной, но в существовании подобных волн в то время уже никто не сомневался.

Мы вернемся к вопросу об этих «волнах» немного позже. А пока воспользуемся гипотезой де Бройля для объяснения соотношений неопределенностей. Предположим, что на пути строго параллельного пучка электронов с импульсом p поставлен экран с очень узкой щелью, ширина которой в направлении оси x равна d (ось x перпендикулярна к направлению пучка) (рис. 5.7). При про-

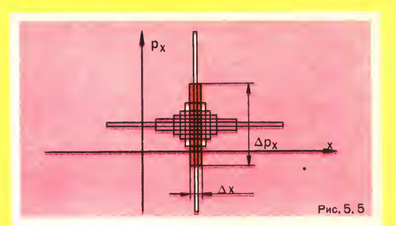


Рис. 5. 5



Рис.5.6

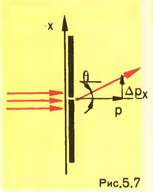


Рис.5.7

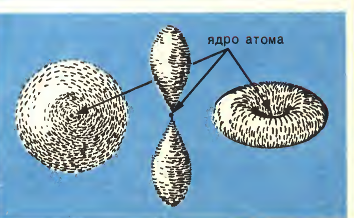


Рис.5.8

хождении электронов сквозь щель происходит дифракция. Согласно классической волновой теории, угол между направлением исходного пучка и направлением на первый дифракционный максимум есть $\theta \approx \lambda/d$. Если рассматривать λ как волновую характеристику электрона и воспользоваться вторым соотношением (5.4), то можно представить угол θ в виде: $\theta \approx \hbar/pd$. Однако как понимать на «языке корпускулярных величин» сам факт существования угла θ? Этот факт означает, что при прохождении через щель электрон приобретает некий импульс Δp_x в направлении оси x. Ясно, что $\Delta p_x \approx p\theta$. Учитывая $\theta \approx \hbar/pd$, получаем: $\Delta p_x d \approx \hbar$. Рассматривая величину d как неопределенность Δx x-координаты электрона, проходящего сквозь щель, приходим к соотношению неопределенностей (5.3).

Волновая функция. Пусть микрообъект находится в состоянии, где x-проекция его импульса имеет определенное значение (значение p_0). Мы уже знаем, что в этом состоянии x-координата микрообъекта имеет сколь угодно большую неопределенность. Иными словами, микрообъект может быть обнаружен где угодно на оси x.

Означает ли все это, что в данном случае мы ничего не можем утверждать об x-координате микрообъекта? Нет, не означает. Оказывается, мы можем говорить о *вероятности* того, что x-координата микрообъекта окажется в некотором интервале значений от x до $x+\Delta x$. Эту вероятность записывают так: $|\Psi_{p_0}(x)|^2\Delta x$.

Мы видим, что плотность вероятности обнаружить рассматриваемый микробъект в точке x записывается в виде квадрата модуля некоторой функции $\Psi_{p_0}(x)$. Эту функцию принято называть *волновой функцией.* Читатель не должен придавать буквального смысла термину «волновая». Дело в том, что в 30-х годах исследователи микромира чрезмерно увлекались волновыми представлениями (под влиянием опытов по дифракции электронов). В те времена говорили не о «квантовой механике», а о «волновой механике».

Итак, состояние микрообъекта, в котором x-проекция импульса имеет значение p_0, а x-координата не имеет определенного значения, описывается волновой функцией $\Psi_{p_0}(x)$, квадрат модуля которой есть плотность вероятности того, что рассматриваемый микрообъект будет обнаружен в точке x. Подчеркнем, что результат измерения координаты микрообъекта в состоянии $\Psi_{p_0}(x)$ оказывается всякий раз *случайным.* То или иное значение координаты реализуется с плотностью вероятности $|\Psi_{p_0}(x)|^2$.

Мы выбрали лишь одно состояние микрообъекта, не касаясь, например, состояний, в которых и импульс, и координата характеризуются неопределенностью. Кроме того, мы ограничились координатой и импульсом, не касаясь других величин, например энергии или момента импульса. Полагаем, что этого достаточно для того, чтобы понять главную мысль: всякое состояние микрообъекта описывается функцией, определяющей вероятность (или плотность вероятности) каких-то характеристик микрообъекта. Отсюда видно, что квантовая

механика уже одного микрообъекта является *вероятностной* теорией.

Электрон в атоме. Электроны в атоме могут находиться в разных состояниях. Изменение состояния электрона может быть, в частности, связано с переходом атома с одного уровня энергии на другой. Будем записывать возможные состояния электрона в некотором атоме при помощи волновых функций $\Psi_j(x, y, z)$, где j — набор некоторых чисел, характеризующих то или иное состояние; x, y, z — координаты электрона. В соответствии со сказанным выше, заключаем, что $|\Psi_j(x, y, z)|^2$ есть плотность вероятности обнаружить в точке (x, y, z) электрон, находящийся в состоянии j. Теперь представим себе «объект», плотность которого в разных точках пространства пропорциональна $|\Psi_j(x, y, z)|^2$. Можно вообразить своеобразное облако с меняющейся от точки к точке плотностью. Внутри облака плотность наибольшая. По мере приближения к поверхности облака она спадает до нуля, в результате чего выявляется некоторая *форма* облака (хотя и без четкой ограничивающей поверхности).

Вот такое «облако» и есть вероятностный «образ» электрона в атоме. На рисунке 5.8 показаны несколько «электронных облаков» — для нескольких состояний электрона в атоме. Подобные картинки пришли на смену устаревшим представлениям об электроне, движущемся внутри атома по орбите.

Сложение амплитуд вероятностей и интерференция

В этом параграфе мы убедимся, что в микромире вероятности подчиняются законам, с которыми мы ранее не встречались. Примечательно, что эти специфические законы позволяют сделать довольно неожиданный вывод: интерференция и дифракция в принципе возможны и в отсутствие каких бы то ни было волн. Они могут быть следствием *особых правил сложения для вероятностей.*

Загадочное поведение микрообъекта в интерферометре. Не касаясь технических подробностей, рассмотрим опыт, в котором некоторые микрообъекты проходят через своеобразный интерферометр в виде двух близко расположенных щелей и затем регистрируются в той или иной точке специального экрана (рис. 5.9). Будем рассматривать только x-координату зарегистрированных микрообъектов. Чтобы в дальнейшем иметь дело не с *плотностью вероятности*, а с самой *вероятностью*, предположим, что ось x на экране разбита на одинаковые маленькие участки, так что, говоря о вероятности попасть в точку x, мы будем подразумевать вероятность попадания в соответствующий участок оси вблизи точки x.

Предположим, что щель A закрыта, а щель B открыта. Зарегистрировав достаточно большое число микрообъектов, мы получим на экране некоторое распределение, описываемое функцией $w_B(x)$ (рис. 5.9, *а*). Эта функция есть вероятность микрообъекту, прошедшему через щель B (при закрытой щели A), попасть в точку x. В соответствии с замечаниями, сделанными в предыдущем

параграфе,

$$w_B(x) = |\Psi_B(x)|^2, \qquad (5.5)$$

где $\Psi_B(x)$ — волновая функция, описывающая микрообъект, прошедший через щель B.

Заметим, что в последнее время термин «волновая функция» все чаще заменяют более подходящим термином — «амплитуда вероятности» (или «амплитуда плотности вероятности»). Тем самым подчеркивается вероятностный характер описания состояния микрообъекта. Ниже мы будем говорить только об *амплитуде вероятности*. Итак, $\Psi_B(x)$ есть амплитуда вероятности попадания в точку x для микрообъекта, прошедшего через щель B (при закрытой щели A).

Предположим теперь, что закрыта щель B, а щель A открыта. В этом случае на экране реализуется распределение $w_A(x)$ (рис. 5.9, *б*):

$$w_A(x) = |\Psi_A(x)|^2, \qquad (5.6)$$

где $\Psi_A(x)$ — амплитуда вероятности попадания в точку x для микрообъекта, прошедшего через щель A (при закрытой щели B).

Наконец, откроем обе щели. Естественно считать, что проходя через одну из щелей, микрообъект «не ощущает» другой щели. Можно сказать, что ему «безразлично», открыта или закрыта эта другая щель. Но в таком случае распределение на экране должно быть суммой распределений (5.5) и (5.6), что, кстати, отвечает правилу *сложения вероятностей:*

$$w_{AB}(x) = w_A(x) + w_B(x) = |\Psi_A(x)|^2 + |\Psi_B(x)|^2. \qquad (5.7)$$

В действительности же на экране наблюдается не распределение (5.7), а типичное *интерференционное* распределение (рис. 5.9, *в*). Получается, что, проходя через одну щель, микрообъект каким-то образом «ощущает» другую щель. Или же, что столь же непонятно, микрообъект умудряется как-то пройти сразу через обе щели. Как же в действительности он проходит через интерферометр?

«Подглядывание» уничтожает интерференционную картину. Попробуем «подглядеть», как ведет себя микрообъект при обеих открытых щелях. В принципе это вполне осуществимо. Можно, например, поместить вблизи каждой щели источник света и регистрировать фотоны, рассеянные микрообъектом вблизи соответствующей щели. Подобные опыты ставились. Они показали, что всякий раз микрообъект проходит только через одну какую-то щель. При этом оказалось, что распределение на экране описывается функцией (5.7). Это означает, что при «подглядывании» можно выяснить подробности прохождения микрообъекта через интерферометр, но при этом, как оказывается, *уничтожается интерференционное распределение.*

Мы приходим, таким образом, к любопытной ситуации. Если свет выключен («подглядывания» нет), то интерференция наблюдается. При этом неизвестен механизм прохождения микрообъекта

через интерферометр. Если свет включен, то указанный механизм выявляется, но при этом уничтожается интерференция.

Когда надо складывать вероятности, а когда амплитуды вероятностей. Приступим к объяснению отмеченных выше удивительных результатов. Отметим, что в данном случае у микрообъекта имеются две возможности (две альтернативы): либо пройти через щель A, либо через щель B. Если свет выключен, то обе эти альтернативы являются *неразличимыми*. Они становятся *различимыми*, если включить свет и тем самым осуществлять «подглядывание», или, переходя на серьезный язык, наблюдение.

Один из основных выводов квантовой механики гласит: *если альтернативы различимы, то соответствующие им вероятности складываются; если же альтернативы неразличимы, то складываются не сами вероятности, а амплитуды вероятностей.* Следовательно, при включенном свете надо складывать вероятности, а при выключенном — амплитуды вероятностей. В первом случае мы приходим к распределению (5.7); во втором случае получаем распределение

$$w(x) = |\Psi_A(x) + \Psi_B(x)|^2. \qquad (5.8)$$

Это распределение имеет интерференционный характер. Можно было бы показать, что

$$|\Psi_A + \Psi_B|^2 = |\Psi_A|^2 + |\Psi_B|^2 + \left[\frac{\Psi_A}{\Psi_B}|\Psi_B|^2 + \frac{\Psi_B}{\Psi_A}|\Psi_A|^2\right]. \qquad (5.9)$$

Выражение, стоящее здесь в квадратных скобках, как раз и «отвечает» за интерференционный характер распределения $w(x)$. В классической физике вопрос о различимых (неразличимых) событиях не возникает. Там всегда события различимы. В микромире ситуация оказывается качественно иной. Здесь мы встречаемся с возможностью полной *неразличимости* тех или иных случайных событий. Такая возможность опирается на принципиальную *тождественность* всех микрообъектов одного и того же типа. Один электрон похож на другой в гораздо большей степени, чем вошедшие в поговорку две капли воды. Конечно, электроны могут находиться в разных состояниях, что позволяет проводить между ними различие. Однако сам по себе электрон (как физический объект) ничем не отличается от любого другого электрона. Мы имеем здесь дело с абсолютной тождественностью. Она в конечном счете и приводит к неразличимым альтернативам.

Мы видим, что явление интерференции не следует ограничивать рамками волновых представлений. Интерференция в микроявлениях не обязательно связана с волнами, она может быть следствием специфических вероятностных законов, а точнее, следствием того, что для неразличимых событий надо складывать не сами вероятности, а их амплитуды.

Квантовомеханическая суперпозиция. Представим

$$\Psi_A(x) + \Psi_B(x) = \Psi(x). \qquad (5.10)$$

Функция $\Psi(x)$ рассматривается в квантовой механике наравне

с функциями $\Psi_A(x)$ и $\Psi_B(x)$. Как и они, эта функция описывает некоторое состояние, или, иными словами, как и Ψ_A, и Ψ_B; это есть амплитуда вероятности некоторого случайного события. В данном случае $\Psi(x)$ есть амплитуда вероятности попасть в точку x для микрообъекта, проходящего через интерферометр с двумя открытыми щелями. Об этой амплитуде говорят как о *суперпозиции* амплитуд Ψ_A и Ψ_B.

Наглядно представить себе подобную суперпозицию нельзя. Иначе пришлось бы всерьез полагать, что микрообъект проходит одновременно и через щель A, и через щель B. Попытки же выявить подробности картины немедленно приводят к *разрушению суперпозиции*. Она разрушается всякий раз либо в пользу Ψ_A (данный микрообъект прошел через щель A), либо в пользу Ψ_B (микрообъект прошел через щель B).

Мы сталкиваемся здесь с еще одним проявлением случайного. Ранее мы уже отмечали, что попадание микрообъекта в то или иное место экрана есть случайное событие; вероятности (5.7) и (5.8) как раз и характеризуют подобные случайные события. Оказывается, что случаен также и «выбор» микрообъектом той или иной щели. Микрообъект проходит через щель A с вероятностью, пропорциональной $|\Psi_A|^2$, а через щель B с вероятностью, пропорциональной $|\Psi_B|^2$.

Волна или сложение амплитуд вероятностей? Волна как нельзя лучше объясняет возникновение интерференционной картины. Но волновые представления не могут объяснить обратного явления — уничтожение интерференционной картины при «подглядывании». Иными словами, волна может объяснить *возникновение* квантовомеханической суперпозиции, но не может объяснить *разрушение* суперпозиции в процессе наблюдения.

Убедившись в этом, а также в бесплодности попыток наделить «волны де Бройля» какой-либо материальной сущностью, физики признали, что эти «волны» не имеют ничего общего с реально существующими волнами. Недаром возникло очень выразительное название — *волны вероятности*. Постепенно термин «волновая механика» был повсеместно заменен термином «квантовая механика», а «волновую функцию» все чаще стали называть «амплитудой вероятности».

Таким образом, мы должны объяснять интерференцию и дифракцию микрообъектов не какими-то загадочными волнами, а необходимостью складывать не вероятности, а амплитуды вероятностей в тех случаях, когда рассматриваемые альтернативы являются неразличимыми. Вероятностный подход исчерпывающе объясняет как возникновение квантовомеханической суперпозиции, так и ее разрушение.

В заключение рассмотрим один пример, позволяющий ясно почувствовать ограниченность волнового подхода. Речь пойдет о рассеянии очень медленных нейтронов на кристалле.

Рассеяние нейтронов на кристалле. Пучок нейтронов, имеющих энергию порядка всего лишь 0,1 эВ, рассеивается на кристалле.

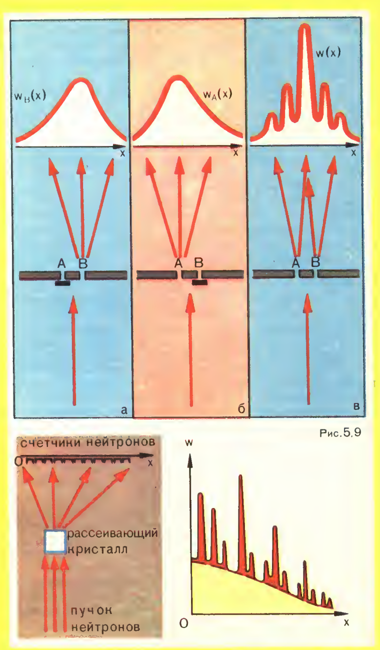


Рис. 5.9

Рис. 5.10

Рис. 5.11

Рассеянные ядрами кристалла нейтроны регистрируются линейкой детекторов, расположенных вдоль оси x (рис. 5.10). Кристаллический образец содержит N ядер. Имеется, следовательно, N альтернатив. Каждая альтернатива отвечает рассеянию нейтрона на том или ином ядре. Обозначим через $\Psi_j(x)$ амплитуду вероятности попасть в детектор в точке x для нейтрона, рассеянного на j-м ядре.

Любопытно, что рассеяние нейтрона на том или ином ядре может происходить двумя способами. В одном случае рассеяние сопровождается *переворачиванием спина* нейтрона, а в другом — не сопровождается. Поясним, что следует понимать под этими словами. Нейтрон можно условно представить в виде своеобразного вращающегося волчка. Волчок может вращаться либо в одну, либо в другую сторону. Соответственно говорят о спине нейтрона, направленном либо вверх, либо вниз. Ядра кристалла также напоминают вращающиеся волчки, т. е. им можно приписать какое-то направление спина. При столкновении с ядром нейтрон-волчок может либо не изменить, либо изменить направление своего вращения. В первом случае спин нейтрона остается неизменным, а во втором — переворачивается. Если при рассеянии нейтрон изменил направление своего вращения, то как-то изменится и направление вращения того самого ядра, на котором произошел данный акт рассеяния. Следовательно, если происходит рассеяние с переворачиванием спина нейтрона, то в этом случае мы имеем дело с *различимой* альтернативой. Мы можем утверждать, что рассеяние произошло именно на том ядре, у которого изменилось направление вращения. Если же рассеяние произошло без переворачивания спина, то в этом случае принципиально невозможно указать, на каком ядре рассеялся нейтрон; здесь мы имеем дело с *неразличимыми* альтернативами.

Пусть φ — амплитуда вероятности рассеяться с переворачиванием спина, а χ — без переворачивания. Обозначим через $\Phi(x)$ амплитуду вероятности попасть в x нейтрону с перевернувшимся спином, а через $X(x)$ — то же для нейтрона с неперевернувшимся спином. Фиксируемое детекторами распределение рассеянных нейтронов можно записать в виде:

$$w(x)=|\varphi|^2|\Phi(x)|^2+|\chi|^2|X(x)|^2. \tag{5.11}$$

Альтернативы, отвечающие разным типам рассеяния нейтрона, естественно, различимы; поэтому вероятность $w(x)$ состоит из двух слагаемых (складываются две вероятности). В свою очередь, каждое слагаемое есть произведение двух вероятностей.

Далее выразим $|\Phi(x)|^2$ и $|X(x)|^2$ через амплитуды $\Psi_j(x)$. Если нейтрон рассеивается с переворачиванием спина, то альтернативы различимы; поэтому *складываются вероятности* и, следовательно,

$$|\Phi(x)|^2=\sum_{j=1}^{N}|\Psi_j(x)|^2. \tag{5.12}$$

Если же при рассеянии спин нейтрона не переворачивается, то

альтернативы неразличимы; поэтому *складываются амплитуды вероятностей* (возникает суперпозиция амплитуд) и, следовательно,

$$|X(x)|^2 = \left| \sum_{j=1}^{N} \Psi_j(x) \right|^2. \tag{5.13}$$

Подставляя (5.12) и (5.13) в (5.11), получаем:

$$w(x) = \left[|\varphi|^2 \sum_{j=1}^{N} |\Psi_j(x)|^2 \right] + \left[|\chi|^2 \left| \sum_{j=1}^{N} \Psi_j(x) \right|^2 \right]. \tag{5.14}$$

На рисунке 5.11 дано получаемое в эксперименте распределение рассеянных нейтронов $w(x)$. Оно состоит из плавно изменяющегося «фона» и набора интерференционных максимумов. «Фон» описывается в (5.14) слагаемым в первых квадратных скобках, а интерференционные максимумы — слагаемым во вторых квадратных скобках.

Если использовать волновые представления, то придется предполагать, что при рассеянии без переворачивания спина нейтрон проявляет свойства волны (возникает интерференционная картина), тогда как при рассеянии с переворачиванием спина нейтрон не проявляет волновых свойств (интерференционная картина не возникает). Искусственность такого предположения вполне очевидна.

Вероятность и причинность

Читатель. По-моему, в микромире чересчур много случайного. Совершенно случайно, без какого-либо воздействия на него нейтрон вдруг превращается в три новые частицы. Какой-нибудь атом пребывает в покое много лет и вдруг, ни с того ни с сего, взрывается, превращаясь в атом другого химического элемента. Электрон случайно проходит через такую-то щель в интерферометре и опять же случайно оказывается в некоторой точке экрана. Не означает ли все это, что в явлениях микромира фактически отсутствует *причинность*?

Автор. Нет, не означает. В явлениях микромира особенно ярко проявляется *диалектическое единство случайного и необходимого*. Нейтроны распадаются случайным образом, но их количество изменяется со временем по определенному закону. Электрон случайно попадает в ту или иную точку экрана, однако распределение попаданий многих электронов является необходимым. Нет никаких оснований сомневаться в существовании причинности в микромире. Надо иметь в виду, что причинность в микромире проявляется иначе, чем в макромире. В квантовой механике причинно связаны не сами отдельные *реализовавшиеся* события, а лишь потенциальные возможности реализации этих событий, или, иначе говоря, *вероятности* событий. Амплитуда вероятностей (волновая функция) подчиняется определенному уравнению движения. Зная амплитуду вероятности в начальный момент и используя это уравнение (его называют *уравнением Шредингера*), можно найти амплитуду вероятности в произвольный момент времени.

Читатель. Мне непонятно, почему нейтрон вдруг распадается. Может быть, микрообъекты являются в действительности какими-то сложными системами, физическая сущность которых нам пока неизвестна?

Автор. Этот вопрос уже возникал в нашей первой беседе (см. с. 5). Я уже говорил, что поиски *скрытых параметров,* которые могли бы объяснить, почему, например, нейтрон распался в данный момент времени, оказались безуспешными. Но мне хотелось бы поговорить о том, что скрывается за поставленным вопросом. Задавая этот вопрос, вы исходили из того, что вероятность в микромире *не объективна,* а связана с нашим незнанием *каких-то деталей.* Думаю, что не только на примере микроявлений, но и на многих приводившихся ранее примерах из нашего обычного макромира вы могли убедиться, что вероятность может быть не только субъективной (связанной с неполнотой знаний), но и объективной. Это очень важно. Ведь только тогда, когда вероятность объективна, можно говорить о первичности, фундаментальности вероятностных закономерностей.

Читатель. Поясните свою мысль.

Автор. Если бы вероятность была связана только с недостатком информации, то тогда она в принципе могла бы быть сведена к динамическим соотношениям, предполагающим однозначные предсказания. Это означало бы, что за вероятностными закономерностями прячутся закономерности динамические. В этом случае можно было бы утверждать, что в природе в конечном счете все жестко взаимосвязано.

Читатель. Но разве любое явление, любое событие не имеет в конечном счете какую-то причину?

Автор. Насчет причины вы, безусловно, правы. Однако почему вы полагаете, что существование объективной вероятности означает отсутствие причинности?

Читатель. Объективная вероятность предполагает и объективную случайность. А такая случайность проявляется без всякой причины. Просто потому, что она есть случайность.

Автор. Я бросаю кубик — выпадает, скажем, четверка. Вы бросаете — выпадает единица. Как вы думаете, выпадение четверки и выпадение единицы — объективно случайные события или нет?

Читатель. Каждое из этих событий имеет определенные причины. Выпадение той или иной цифры зависит в конечном счете от положения кубика в руке, взмаха кисти, толчка, сопротивления воздуха, расстояния от руки до пола.

Автор. Верно. И тем не менее эти события — объективно случайные. Ведь, бросая свой кубик, вы не интересовались тем, как бросал свой я. Мы вообще не интересовались тем, как бросается кубик, не пытались контролировать и как-либо направлять ни собственные действия, ни действия партнера. Поэтому выпадения четверки у меня и единицы у вас — объективно случайные события. Выпадение единицы никак не связано с тем, что перед этим выпала четверка.

Читатель. Я не совсем понимаю вашу мысль.
Автор. Приведу другой пример. Рассматриваются события — заказы такси по телефону. За каждым заказом скрывается целая цепь причин. Для диспетчера таксомоторного парка поступающие заказы — объективно случайные события. Причем вовсе не потому, что он не знает упомянутой цепи причин, а вследствие совершенно объективного обстоятельства — отсутствия взаимной согласованности в действиях людей, которые делают заказы. Тут события рассматриваются как бы в двух разных плоскостях. В одной они объективно случайны, в другой — каждое из них имеет определенные причины. Как видите, объективная вероятность прекрасно согласуется с причинностью.
Читатель. Вы привели примеры из человеческой практики. А как же быть с микроявлениями? Возьмем опять пример с распадом нейтрона. Пусть это событие объективно случайно в какой-то «плоскости». А в какой «плоскости» надо искать причины, вызвавшие распад нейтрона?
Автор. Распад нейтрона действительно объективно случаен. Мы принципиально (а не в силу незнания деталей) не можем управлять длительностью «жизни» данного нейтрона. Нет у нейтрона и «внутренних часов». Как уже отмечалось, нейтроны «не стареют». Это проявляется в том, что вероятность нейтрону прожить некоторое время не зависит от того, сколько он уже «прожил» к моменту, от которого начинается отсчет времени. Будучи объективно случайным, распад нейтрона в то же время не есть беспричинно происходящее событие. Замечу, что, когда мы говорим о *самопроизвольном* поведении того или иного микрообъекта, мы допускаем некоторую неточность. Строго говоря, *самопроизвольно* может вести себя только строго *изолированный* объект. И вот тут мы подходим к принципиальному обстоятельству, о котором до сих пор еще не говорилось.
Дело в том, что микрообъект по своей природе *не изолированный объект* — он взаимодействует со всем окружающим миром. Сама его сущность реализуется в том или ином виде в зависимости от условий конкретной обстановки. Термин «взаимодействие» следует при этом понимать шире, чем это понимается при рассмотрении обычных (силовых) взаимодействий.
Читатель. Очередные загадки квантовой механики.
Автор. Дело не в загадках. Просто на определенном уровне изучения физических явлений объекты принципиально утрачивают свою изолированность. Так стираются существовавшие до того четкие грани между полем и веществом. На первый план выдвигаются взаимопревращения частиц. На уровне микромира приобретает особый смысл идея *единства мира и всеобщей связи явлений.*
Читатель. Как можно наглядно представить «неизолированность» распадающегося нейтрона?
Автор. В квантовой теории *вакуум* представляют не как пустоту, а как пространство, где случайным образом рождаются и уничтожаются различные частицы. Нейтрон взаимодействует с ними.

ГЛАВА 6 Вероятность в биологии

Явления случайного порядка, каковыми с первого взгляда в отдельности представляются мутации, идущие в разных направлениях, в конечном итоге выявляют закономерный процесс.

Н. И. Вавилов

В процессе передачи из поколения в поколение генетические программы в результате многих причин изменяются случайно и ненаправленно, и лишь случайно эти изменения оказываются приспособительными.

Б. М. Медников

Интродукция

Жан Батнст Ламарк (1744—1829). В 1809 году вышла в свет «Философия зоологии» французского ученого Жана Батиста Ламарка. В этом труде была предпринята первая попытка создания теории эволюции видов. Попытка оказалась неудачной. Создавая свою теорию, Ламарк исходил из двух ошибочных представлений. Во-первых, он полагал, что во всех живых существах заложено внутреннее стремление к совершенствованию. В этом он усматривал движущую силу эволюции. Разумеется, никакой таинственной внутренней силы, заставляющей все виды эволюционировать в направлении прогресса, не существует. Да и откуда бы она могла взяться? Разве лишь благодаря вмешательству «творца». Ясно, что подобная точка зрения приводит в конечном итоге к вере в бога.

Во-вторых, Ламарк считал, что внешняя среда непосредственным образом влияет на изменение формы тех или иных органов живых существ. Когда-то существовали жирафы с короткой шеей. По каким-то причинам изменились условия их обитания. Пища оказалась высоко над поверхностью (листва высоких деревьев). Чтобы добраться до пищи, жирафам приходилось все время тянуть кверху шеи. Это происходило из поколения в поколение. В результате длительных упражнений шея жирафов вытянулась.

В качестве доказательства Ламарк приводил общеизвестный факт превращения физически слабых людей в атлетов в результате регулярных занятий спортом. Он сформулировал следующий закон: «У каждого животного, не завершившего еще своего развития, более частое и продолжительное употребление какого-нибудь органа укрепляет этот орган, развивает его, увеличивает и придает ему силу, пропорциональную продолжительности употребления, тогда как постоянное отсутствие употребления какого-либо органа постепенно его ослабляет, приводит к упадку, непрерывно уменьшает его способности и, наконец, заставляет его исчезнуть».

Ламарк глубоко ошибался. Известно, что натренированные мышцы, равно как и приобретенные навыки, по наследству не передаются. Используя современную терминологию, мы можем сказать, что Ламарк не понимал различия между фенотипом и генотипом. *Генотип* — это своего рода наследственная конституция организма, совокупность наследственных зачатков, которую организм получил от родителей *Фенотип* — совокупность внешних и внутренних признаков рассматриваемого организма; сюда входят все наблюдаемые признаки — анатомические, физиологические, психические и др. Фенотип изменяется в течение жизни организма в результате взаимодействия между генотипом и окружающей средой. Регулярными занятиями гимнастикой, упорной учебой, правильной организацией труда и отдыха каждый из нас может улучшить свой фенотип. Все это однако не влияет на генотип.

Чарлз Дарвин (1809—1882). Правильная эволюционная теория была создана великим английским ученым Чарлзом Дарвином. Эта теория получила название дарвинизма. Она была изложена в книге «Происхождение видов путем естественного отбора», вышедшей в 1859 году.

Учение Дарвина опирается на три фактора: *изменчивость, наследственность, естественный отбор.* Внешняя среда, воздействуя на организм, может приводить, в частности, к случайным изменениям генотипа. Эти изменения передаются по наследству и постепенно накапливаются в потомстве. Характер изменений различен. Некоторые случайно оказываются более благоприятными с точки зрения приспособления организмов к условиям внешней среды, другие менее благоприятными, третьи вообще вредными. При накоплении в потомстве тех или иных случайных изменений начинает проявляться действие естественного отбора. Организмы, оказавшиеся менее приспособленными, дают меньшее потомство, преждевременно погибают; в конечном счете их вытесняют более приспособленные.

Описывая сущность учения Дарвина, мы специально подчеркнули важную роль случайностей. Читатель должен узнать знакомую идею *отбора информации из шума.*

Рассматривая эволюцию видов, Ламарк признавал, по сути дела, лишь *голую необходимость.* Изменились условия внешней среды — и организм за счет упражнения-неупражнения соответствующих органов необходимым образом видоизменяется. Такая «эволюция» с необходимостью идет только в направлении усложнения организации организмов, как если бы в каждом виде действительно было заложено внутреннее стремление к прогрессу.

Дарвин же рассматривал эволюцию с позиций диалектического единства необходимого и случайного. Безразличная природа вызывает в организмах случайные наследственные изменения, затем через естественный отбор безжалостно отсекает тех, кто случайно оказался менее приспособленным, и оставляет тех, кто случайно оказался достаточно приспособленным к условиям внешней среды. В результате с необходимостью совершается процесс эволю-

ционного развития видов. Развитие идет по пути отбора более приспособленных, при этом природе безразлично, будут ли эти организмы более сложно или, напротив, менее сложно организованы. Возможности приспособления в тех или иных условиях могут быть весьма разнообразны. В итоге и возникает наблюдаемое нами многообразие видов животных и растений. Как известно, на Земле сейчас имеются около 1,5 миллионов видов животных и около 0,5 миллиона видов растений.

Учение Дарвина получило всеобщее признание. Однако в нем есть одно «больное место», на которое указал Дарвину в 1867 году преподаватель из Эдинбурга Флеминг Дженкинс. Он заметил, что в дарвинской теории нет ясности в вопросе о том, как осуществляется накопление в потомстве тех или иных изменений. Ведь сначала изменения признака происходят лишь у некоторых особей. Эти особи скрещиваются с нормальными особями. В результате, утверждал Дженкинс, должно наблюдаться не *накопление* измененного признака в потомстве, а, напротив, его *разбавление*, постепенное стирание — вплоть до исчезновения (в первом поколении потомства остается 1/2 изменения, во втором поколении 1/4 изменения, в третьем 1/8 изменения, в четвертом 1/16 изменения и т. д.).

В течение пятнадцати лет, до самой своей кончины, Дарвин размышлял над вопросом, поставленным Дженкинсом. Решения проблемы он так и не нашел.

А между тем это решение существовало уже в 1865 году. Его получил преподаватель монастырской школы в Брюнне (теперь Брно, Чехословакия) Грегор Иоганн Мендель. Увы, Дарвин ничего не знал об исследованиях Менделя. Он так никогда и не узнал о них.

Грегор Иоганн Мендель (1822—1884). Свои знаменитые опыты с горохом Мендель начал проводить за три года до выхода в свет «Происхождения видов». Когда появилась книга Дарвина, он внимательно прочитал ее и в дальнейшем живо интересовался всеми работами Дарвина. Говорят, что однажды Мендель заметил по поводу дарвинской теории: «Это еще не все, еще чего-то здесь не хватает». Исследования Менделя как раз и были направлены на то, чтобы заделать «брешь» в теории Дарвина. Мендель занимался гибридизацией, он хотел проследить судьбу изменений генотипов в разных поколениях гибридов. Объектом исследования Мендель выбрал горох.

Мендель взял два сорта гороха — с желтыми и с зелеными семенами. Скрестив эти два сорта, он обнаружил в первом поколении гибридов горох только с желтыми семенами. Зеленый горох словно сквозь землю провалился. Затем Мендель произвел самоопыление полученных гибридов и получил второе поколение гибридов. В этом поколении снова появились особи с зелеными семенами. Правда, их оказалось заметно меньше, чем с желтыми. Мендель тщательно подсчитал число тех и других и получил, что число особей с желтыми семенами относится к числу особей с

зелеными семенами как

$x:y=6022:2001=3{,}01:1.$

Параллельно Мендель проводил еще шесть опытов. В каждом опыте он использовал два сорта гороха, различавшихся по какому-либо одному определенному признаку. Так, в одном из опытов он скрестил горох с гладкими семенами и горох с морщинистыми семенами. В первом поколении гибридов он наблюдал только растения с гладкими семенами. Во втором поколении появились также растения с морщинистыми семенами. Отношение числа особей с гладкими семенами к числу особей с морщинистыми семенами составило

$x:y=5474:1850=2{,}96:1.$

В остальных пяти опытах скрещивались сорта, различающиеся либо по окраске кожуры, либо по форме плода, либо по его окраске в незрелом состоянии, либо по расположению цветков, либо по размерам растений (карлики и гиганты).

В каждом опыте в первом поколении гибридов проявлялся только один из двух противоположных родительских признаков. Мендель назвал этот признак *доминантным*. Другой признак, тот, который временно исчезал, он назвал *рецессивным*. В первом из рассмотренных выше опытов доминантным признаком был желтый цвет семян, а рецессивным — зеленый цвет. Во втором опыте доминантный признак — гладкие семена, рецессивный — морщинистые семена. Отношение $x:y$, т. е. числа особей с доминантным признаком к числу особей с рецессивным признаком среди гибридов второго поколения для этих двух опытов, мы уже приводили. В остальных пяти опытах Мендель получил:

$x:y=705:224=3{,}15:1;$
$x:y=882:299=2{,}95:1;$
$x:y=428:152=2{,}82:1;$
$x:y=651:207=3{,}14:1;$
$x:y=787:277=2{,}84:1.$

Во всех случаях отношение $x:y$ оказывается достаточно близким к отношению $3:1$.

В итоге Мендель мог с уверенностью утверждать: при скрещивании растений с противоположными признаками происходит не разбавление признаков (как полагал Дженкинс), а *подавление одного признака другим*, в связи с этим необходимо различать доминантные и рецессивные признаки;

в гибридах первого поколения проявляется только доминантный признак, рецессивный признак полностью подавлен (правило *единообразия гибридов первого поколения*);

гибриды первого поколения при размножении самоопылением *расщепляются:* во втором поколении появляются особи как с доминантным, так и с рецессивным признаками, причем отношение числа первых к числу вторых равно примерно $3:1$.

Мендель, однако, не остановился на этом. Он произвел самоопыление гибридов второго поколения и получил гибриды третьего, а затем и четвертого поколения. Ученый обнаружил, что гибриды второго поколения с рецессивным признаком при дальнейшем размножении не расщепляются ни в третьем, ни в четвертом поколениях. Так же ведет себя примерно треть гибридов второго поколения с доминантным признаком. Две трети гибридов второго поколения с доминантным признаком расщепляются при переходе к гибридам третьего поколения, причем опять-таки в отношении 3:1. Получившиеся в результате этого расщепления гибриды третьего поколения с рецессивным признаком и треть гибридов с доминантным признаком при переходе к четвертому поколению не расщепляются, а остальные гибриды третьего поколения расщепляются, причем снова в отношении 3:1.

Заметим, что явление расщепления гибридов демонстрирует важное обстоятельство: особи с одинаковыми внешними признаками могут обладать разными наследственными свойствами, что и обнаруживается во внешних признаках их потомства. Мы видим, что по фенотипу нельзя судить с достаточной полнотой о генотипе. Если особь не обнаруживает в потомстве расщепления, то ее называют *гомозиготной*; если же при размножении она обнаруживает расщепление, то ее называют *гетерозиготной*. Пример гомозиготных особей — все особи с рецессивным признаком среди гибридов второго поколения.

Полученные Менделем результаты хорошо просматриваются на рисунке 6.1, где желтым цветом показаны организмы с доминантным признаком, а зеленые — с рецессивным. Глядя на этот рисунок, нетрудно уловить определенную закономерность. Мендель разгадал эту закономерность и тем самым раскрыл механизм передачи наследственных признаков от поколения к поколению. Мендель понял, что разгаданная им закономерность имеет вероятностный характер.

Конечно, наблюдения над гибридами производились и до Менделя. Достаточно, например, привести записи современника Менделя Шарля Нодэна, работавшего садовником в Ботаническом саду в Париже: «Начиная со второго поколения, облик гибридов изменяется заметным образом. Столь совершенное единообразие гибридов первого поколения сменяется обычно крайней пестротой форм, одни из которых приближаются к видовому типу отца, другие — матери...» Но до Менделя никто не предпринял систематизированных исследований, с учетом отдельных выделенных признаков, с подсчетом чисел проявлений тех или иных признаков в различных поколениях гибридов. Мендель был первым, кто все это проделал, потратив на опыты восемь лет. Поэтому, в отличие от всех своих предшественников, Мендель понял *закономерности наследственной передачи признаков*.

Здесь уместно сделать передышку, с тем чтобы в следующем параграфе рассмотреть открытые Менделем законы гибридного скрещивания с позиции современной генетики. А пока сообщим лишь,

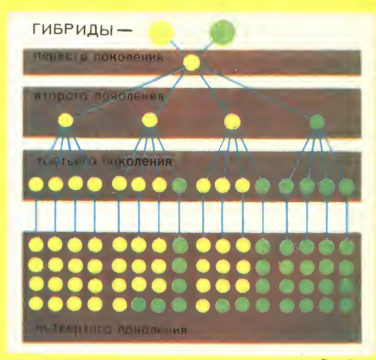


Рис. 6.1

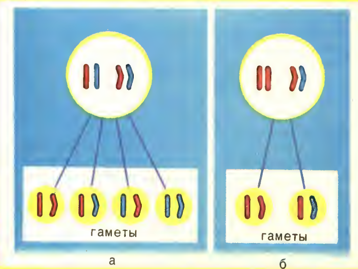


Рис. 6.2

что результаты своих исследований Мендель доложил в феврале 1865 года Обществу естествоиспытателей в Брюнне. Слушатели не поняли исключительной важности представленного доклада. Они не догадались, что этой работе суждено произвести настоящую революцию в науке о наследственности. В 1866 году доклад Менделя был напечатан в Брюннском бюллетене и разослан по списку 120 научным учреждениям разных стран. К сожалению, Дарвин этого бюллетеня не получил.

Мир давно признал Менделя как основателя современной генетики. Это признание пришло лишь в 1900 году, через пятнадцать лет после кончины талантливого исследователя.

Закономерности случайного комбинирования генов при скрещивании

Хромосомы и гены. Напомним читателю некоторые сведения из *цитологии* — раздела биологии, изучающего клетку. Различают два типа клеток — *половые* клетки (*гаметы*) и *неполовые*, или, иначе, *соматические*. В *ядре* каждой клетки находятся нитевидные *хромосомы*, представляющие собой гигантские молекулы дезоксирибонуклеиновой кислоты (сокращенно: ДНК) в соединении с молекулами белков. В хромосомах, а точнее, в молекулах ДНК содержится вся информация, определяющая генотип данного организма. Отдельные участки хромосомы, «ответственные» за те или иные наследственные признаки, называют *генами*. Каждая хромосома содержит несколько сотен генов. Иногда хромосому упрощенно представляют в виде своеобразной нити, на которую, словно бусинки, нанизаны различные гены.

Каждому виду соответствует *определенный набор хромосом*, определяемый количеством хромосом и их генными характеристиками. Например, у овса имеются 42 хромосомы, у плодовой мушки дрозофилы 8 хромосом, у шимпанзе 48 хромосом, у человека 46 хромосом. Ядро каждой соматической клетки содержит полный набор хромосом, соответствующий данному виду. Это означает, что в *каждой клетке* организма содержится вся *наследственная информация*.

Приведенные выше для нескольких видов числа хромосом характеризуют хромосомные наборы в соматических, но не в половых клетках. Каждая половая клетка (гамета) имеет в два раза меньше хромосом, чем соматическая.

Начнем с хромосомного набора соматической клетки. В этот набор входят две *половые хромосомы*. У женских особей обе половые хромосомы одинаковые (две *X-хромосомы*), у мужских особей половые хромосомы разные (одна *X-хромосома* и одна *Y-хромосома*). Неполовые хромосомы, имеющиеся в соматической клетке, разбиваются на *пары*; попавшие в одну пару хромосомы (их называют *гомологичными*) очень похожи друг на друга. Каждая содержит одно и то же число генов, одинаковым образом расположенных в той и другой хромосомных нитях, а главное, отвечаю-

щих за одни и те же виды признаков. Например, у гороха есть пара гомологичных хромосом, каждая из которых содержит ген окраски семян. У этого гена, как и у других, есть две разновидности (их называют *аллелями*) — доминантная и рецессивная. Доминантная разновидность гена окраски (*доминантный аллель*) соответствует желтому цвету, а рецессивная (*рецессивный аллель*) зеленому. Если в обеих гомологичных хромосомах рассматриваемый ген представлен одинаковыми аллелями, то данная особь *гомозиготна* по рассматриваемому признаку. Если же в одной хромосоме содержится один аллель, а в другой гомологичной хромосоме другой, то данная особь *гетерозиготна*. В ее фенотипе проявляется признак, отвечающий доминантному аллелю.

Теперь рассмотрим хромосомный набор гаметы (половой клетки). Гамета имеет только одну половую хромосому. У женской особи — это всегда *X*-хромосома. У мужской особи это может быть либо *X*-хромосома (в одних гаметах), либо *Y*-хромосома (в других гаметах). Кроме единичной половой хромосомы, гамета содержит по одной хромосоме из каждой пары гомологичных хромосом. Допустим, что имеются всего две пары гомологичных хромосом и с каждой парой сопоставляется некоторый определенный признак. Пусть данная особь гетерозиготна по обоим видам признаков. Такая особь будет иметь четыре типа гамет, что хорошо видно из рисунка 6.2, *а* (красным цветом на рисунке показаны хромосомы, несущие доминантные аллели, а синим — рецессивные). В случае, изображенном на рисунке 6.2, *б*, рассматриваемая особь гомозиготна по одному признаку и гетерозиготна по другому. В этом случае имеются только два типа гамет.

При оплодотворении мужская гамета сливается с женской. Оплодотворенная женская гамета (ее называют *зиготой*) имеет полный хромосомный набор. В каждой паре гомологичных хромосом одна хромосома получена от отца, а другая от матери. Организм развивается из зиготы посредством клеточных делений. В каждом случае делению клетки предшествует *дублирование (удвоение)* всех хромосом, содержащихся в ядре клетки. В результате ядро каждой соматической клетки организма содержит тот же самый набор хромосом и генов, какой имела зигота. Когда организм достигает полового созревания, в нем происходят особые процессы, приводящие к образованию гамет. Мы остановимся на этих процессах позднее.

Закон расщепления. Будем рассматривать какой-нибудь один признак. Пусть это будет, как в одном из опытов Менделя, цвет семян гороха. Рассмотрим результаты этого опыта, используя представления современной цитологии.

В первом поколении гибридов все особи гетерозиготны по рассматриваемому признаку. В каждой соматической клетке присутствуют оба аллеля окраски семян — желтый (доминантный аллель) и зеленый (рецессивный). Все семена этих гибридов, естественно, желтые. По рассматриваемому здесь признаку каждый гибрид первого поколения имеет два типа гамет — с доминантным аллелем

(*A*-гаметы) и с рецессивным (*a*-гаметы). Ясно, что существуют как женские, так и мужские *A*-гаметы и *a*-гаметы.

Перейдем к гибридам второго поколения. Каждый новый организм развивается из зиготы, которая образуется при соединении мужской гаметы типа *A* или *a* с женской гаметой типа *A* или *a*. Возможны, очевидно, четыре альтернативы (рис. 6.3):

AA — мужская *A*-гамета соединяется с женской *A*-гаметой,
Aa — мужская *A*-гамета соединяется с женской *a*-гаметой,
aA — мужская *a*-гамета соединяется с женской *A*-гаметой,
aa — мужская *a*-гамета соединяется с женской *a*-гаметой.

Все эти альтернативы равновероятны. Следовательно, среди достаточно большого числа зигот одну четверть будут составлять *AA*-зиготы, одну четверть *aa*-зиготы и, наконец, половину *Aa*-зиготы (здесь объединены варианты *Aa* и *aA* как равноправные с точки зрения наследования признаков). Если зигота содержит хотя бы один доминантный аллель, то в фенотипе проявится доминантный признак (желтый цвет семян). Следовательно, растения, развившиеся из *AA*- и *Aa*-зигот, будут иметь желтые семена, а растения, развившиеся из *aa*-зигот,— зеленые. Мы видим, таким образом, что *вероятность появления óсоби с доминантным признаком равна 3/4, а вероятность появления óсоби с рецессивным признаком равна 1/4.* Отсюда следует полученное Менделем соотношение 3 : 1, количественно характеризующее расщепление признака при переходе от первого поколения гибридов ко второму. Мендель не только выявил это соотношение, но и правильно объяснил его, используя понятие *вероятности.* Все это и составило содержание *первого закона Менделя*, известного также как *закон расщепления.*

Подчеркнем: та или иная зигота образуется в результате *случайной встречи* мужской и женской гамет того или иного типа. Большое число подобных случайных встреч с *необходимостью* выявляет определенную закономерность, которую и выражает первый закон Менделя.

Заметим, что из *AA*- и *aa*-зигот развиваются гомозиготные (по рассматриваемому признаку) особи, тогда как из *Aa*-зигот развиваются гетерозиготные особи, у которых расщепление признака при переходе к следующему поколению будет происходить опять-таки по закону 3 : 1.

Закон независимого распределения генов. Рассмотрим гибриды второго поколения, учитывая теперь не один какой-нибудь признак, а сразу два признака. Будем полагать (это очень важно), что отвечающие за выбранные признаки гены находятся в разных парах гомологичных хромосом. Примером могут служить цвет семян гороха (один признак) и форма семян (другой признак). Будем обозначать: *A* — доминантный аллель цвета (желтый цвет), *a* — рецессивный аллель цвета (зеленый цвет), *B* — доминантный аллель формы (гладкие семена), *b* — рецессивный аллель формы (морщинистые семена).

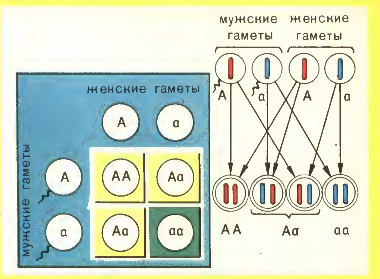

мужские гаметы
женские гаметы
женские гаметы
A
a
мужские гаметы
A
a
AA
Aa
Aa
aa
A
a
A
a
AA
Aa
aa

Рис. 6.3

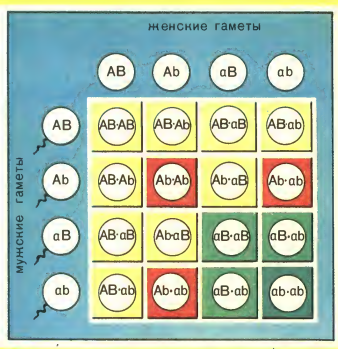

женские гаметы
AB
Ab
aB
ab
мужские гаметы
AB
Ab
aB
ab
AB·AB
AB·Ab
AB·aB
AB·ab
AB·Ab
Ab·Ab
Ab·aB
Ab·ab
AB·aB
Ab·aB
aB·aB
aB·ab
AB·ab
Ab·ab
aB·ab
ab·ab

Рис. 6.4

Каждый гибрид первого поколения имеет четыре типа мужских и четыре типа женских гамет: AB, Ab, aB, ab (напомним рисунок 6.2, *а*). Образование зиготы происходит при соединении двух гамет (мужской и женской) любого из указанных четырех типов. Возможны 16 альтернатив; они даны на рисунке 6.4. *Все эти альтернативы равновероятны.* Следовательно, доля числа зигот разного типа (по отношению к общему числу зигот, которое должно быть достаточно большим) такова: для зигот типа $AB \cdot AB$ — $1/16$, $Ab \cdot Ab$ — $1/16$, $aB \cdot aB$ — $1/16$, $ab \cdot ab$ — $1/16$, $AB \cdot Ab$ (с учетом также $Ab \cdot AB$) — $1/8$, $AB \cdot aB$ (с учетом $aB \cdot AB$) — $1/8$, $AB \cdot ab$ (с учетом $ab \cdot AB$) — $1/8$, $Ab \cdot aB$ (с учетом $aB \cdot Ab$) — $1/8$, $Ab \cdot ab$ (с учетом $ab \cdot Ab$) — $1/8$, $aB \cdot ab$ (с учетом $ab \cdot aB$) — $1/8$. Принимая во внимание подавление рецессивных аллелей соответствующими доминантными, заключаем, что вероятность появления особи с желтыми гладкими семенами во втором поколении гибридов равна сумме вероятностей образования зигот типа $AB \cdot AB$, $AB \cdot Ab$, $AB \times \times aB$, $AB \cdot ab$, $Ab \cdot aB$, т. е. равна $1/16 + 1/8 + 1/8 + 1/8 + 1/8 = 9/16$. Вероятность появления особи с желтыми морщинистыми семенами равна сумме вероятностей образования зигот типа $Ab \cdot Ab$ и $Ab \cdot ab$, т. е. равна $1/16 + 1/8 = 3/16$. Вероятность появления особи с зелеными гладкими семенами равна сумме вероятностей образования зигот типа $aB \cdot aB$ и $aB \cdot ab$, т. е. равна $1/16 + 1/8 = 3/16$. Наконец, вероятность появления особи с зелеными морщинистыми семенами равна вероятности образования зиготы $ab \cdot ab$, т. е. равна $1/16$. Таким образом, числа различных фенотипов (по двум рассматриваемым признакам) во втором поколении гибридов относятся друг к другу как $9:3:3:1$. Все это и составляет сущность *второго закона Менделя,* согласно которому расщепление по одному признаку идет *независимо* от расщепления по другому.

Закон Моргана. Закон независимого распределения генов справедлив, если рассматриваемые гены входят в разные хромосомы в гамете (и соответственно в разные пары гомологичных хромосом в соматической клетке). Если же гены попадают в *одну и ту же* хромосому, то они должны наследоваться *вместе.* Именно этим и объясняется открытое и исследованное американским биологом Т. Морганом отклонение от второго закона Менделя, наблюдаемое всякий раз, когда рассматриваемые признаки определяются *сцепленными* генами, т. е. генами, входящими в одну и ту же хромосому. Совместное наследование сцепленных генов получило название *закона Моргана.*

Томас Хант Морган (1866—1945) является основателем хромосомной теории наследственности. Используя представления о хромосомах, он не только обосновал законы Менделя, но также указал условия их применимости и, кроме того, получил ряд новых важных результатов. К таким новым результатам следует отнести не только закон Моргана, но и открытое Морганом явление перекреста хромосом.

Явление перекреста хромосом. Исследуя передачу по наследству признаков, определяемых сцепленными генами, Морган обнаружил,

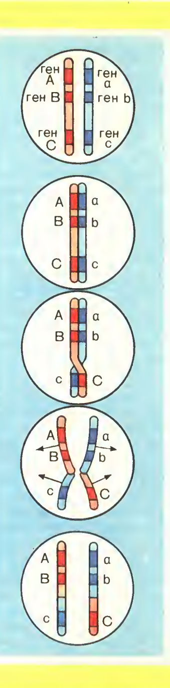


Рис. 6. 5

что сцепление *не является абсолютным:* среди гибридов второго поколения наблюдаются особи, у которых часть сцепленных генов унаследована от одного родителя, а остальные — от другого. Выполнив исследования на плодовой мушке дрозофиле, Морган нашел объяснение этому факту. Он обнаружил, что процесс образования половых клеток в организме (этот процесс называют *мейозом*) начинается со своеобразного «прощального танца» гомологичных хромосом.

Представьте себе две вытянувшиеся гомологичные хромосомные нити, которые, перед тем как разойтись в разные гаметы, тесно прильнули друг к другу (каждый ген к соответствующему гену) и затем несколько раз закрутились вокруг самих себя. Это закручивание хромосом, или, иначе, взаимный перекрест, приводит к тому, что внутриклеточные силы, призванные разъединить хромосомы, оттащить их друг от друга, *разрывают* хромосомы. Место разрыва случайным образом меняется от одной пары перекрещенных хромосом к другой. В результате разрыва в одну гамету отправляется не целая хромосома, а *взаимно дополняющие* друг друга части обеих гомологичных хромосом; другие части этих хромосом отправляются в другую гамету. Этот процесс показан схематически на рисунке 6.5. Подчеркнем, что в момент разрыва соответствующие гены обеих хромосом (речь идет об аллелях) непосредственно контактируют друг с другом. Поэтому, где бы ни произошел разрыв, аллель из одной хромосомы отправится в одну гамету, а аллель из другой хромосомы в другую гамету. Одним словом, не получится так, чтобы в какой-то гамете не оказалось ни одного аллеля рассматриваемого гена. Все это можно представить так, как если бы «танцующие» пары хромосом перед расставанием обменялись друг с другом какими-то частями, причем обязательно соответствующими частями. В конечном счете в каждой образовавшейся гамете все равно окажется полный набор типов генов, присущий данной хромосоме. При этом произойдет *случайное перекомбинирование* отцовских и материнских аллелей.

В явлении перекреста хромосом существенную роль играет случай. *Случайно* место разрыва в той или иной паре хромосом, а следовательно, *случайна* перекомбинация родительских аллелей.

Увеличивая поле действия случайного, явление перекреста хромосом способствует внутривидовому развитию, создавая дополнительные возможности перетасовки родительских генов. В то же время это явление как бы оберегает вид от возможных случайных генетических «посягательств» на него. Допустим, что произошло случайное скрещивание особей двух разных видов и появились гибриды. У этих гибридов в каждой «гомологичной паре» будут объединены хромосомы, весьма отличающиеся одна от другой по своей генной структуре (ведь эти хромосомы взяты от родителей, относящихся к разным видам!). Когда наступит время формирования половых клеток, такие хромосомы не смогут вследствие существенных взаимных различий исполнить совместный «прощальный танец». В результате не смогут образоваться гаметы, а сле-

довательно, и не появятся гибриды второго поколения. Вот почему мулы (гибрид лошади и осла) не имеют потомства.

Мальчик или девочка? Мы уже отмечали, что обе половые хромосомы женщины одинаковы — это *X*-хромосомы. Половые хромосомы мужчины, напротив, различны — *X*-хромосома и *Y*-хромосома. Примерно половина мужских гамет несет *X*-хромосому, другая половина — *Y*-хромосому. Если с женской гаметой соединяется *X*-гамета мужчины, то образуется *XX*-зигота, из нее разовьется девочка. Если же с женской гаметой соединяется *Y*-гамета мужчины, то образуется *XY*-зигота, из нее разовьется мальчик.

Мутации

Мы рассмотрели случайные изменения генетических программ, происходящие при скрещивании в результате комбинирования родительских генов. Все эти изменения ограничены имеющимся фондом генов. Новые гены при этом не создаются. Вместе с тем наблюдаются случайные наследственные изменения, не связанные с комбинированием генов. Они обусловлены действием внешней среды на генную структуру хромосом, а также случайными нарушениями в биологическом механизме, обеспечивающем сохранение генетической информации при делении соматических клеток и при мейозе. Эти наследственные изменения называют *мутациями*.

Некоторые проявления мутаций. Существует серьезное заболевание, проявляющееся в том, что кровь человека утрачивает способность к свертыванию. Это заболевание называют *гемофилией*. Оно передается по наследству и встречается только у мужчин. Выяснено, что гемофилия — следствие *мутации* одного из генов, находящихся в половой *X*-хромосоме. Поскольку у женщины две *X*-хромосомы, то смутировавшему гену в одной из них противостоит нормальный ген в другой. Смутировавший ген рецессивен. Он подавляется нормальным геном. Поэтому женщины и не заболевают гемофилией. Иное дело мужчины. Набор половых хромосом мужчины состоит из двух *разных* хромосом — *X*-хромосомы и *Y*-хромосомы. В данном случае нет пáрного нормального гена, который мог бы подавить ген гемофилии. В результате мужчина, получивший от фенотипически здоровой матери *X*-хромосому со смутировавшим геном, заболевает гемофилией.

К счастью, чаще мутации проявляются более безобидно. Коротко-палая кисть, шестой палец, сердце справа — относительно редкие проявления мутации. Более часто наблюдаются такие мутации, как, например, разный цвет глаз, значительное облысение (включая форму лысины), необычная окраска шерсти у животных и т. д. Относительно часто встречаются мутации у растений. Они выражаются весьма разнообразно, затрагивая формы ствола, листьев, цветков.

Причины появления мутаций. Та или иная мутация — довольно редкое явление. Например, вероятность того, что взятая наугад гамета с *X*-хромосомой будет содержать мутацию, связанную с гемо-

филией, равна всего 10^{-5}. Другие мутации происходят еще реже — в среднем с вероятностью примерно 10^{-6}. Надо, однако, принимать во внимание многообразие мутаций. Они могут затрагивать самые разные гены из огромного их числа, приходящегося на каждую гамету. Надо учитывать также, что мутации передаются по наследству, они *накапливаются*. В итоге мутации оказываются не такими уж редкими событиями. Подсчитано, что примерно среди каждых десяти гамет человека можно обнаружить гамету, несущую какую-нибудь мутацию.

Появление конкретной мутации — *случайное* событие. Но у этого события есть объективные причины. Организм развивается из зиготы в результате многократных делений клеток. Процесс деления клетки начинается с того, что в ее ядре происходит самоудвоение (редупликация) хромосом и, следовательно, молекул ДНК. Каждая молекула ДНК как бы воссоздает свою точную копию — с таким же набором генов. Сложный процесс редупликации молекулы ДНК не обходится без случайных нарушений. Как известно, генетическая информация записывается в ДНК сверхэкономно — на молекулярном уровне. При копировании информации возможны различного рода «опечатки», обусловленные тепловым движением молекул вещества. «Опечатки» возникают вследствие неизбежных *флуктуаций* в поведении частиц вещества. Например, в молекуле ДНК во время ее самоудвоения может случайно возрасти количество ионов водорода вблизи какого-нибудь азотистого основания. Такая флуктуация может привести к отщеплению данного основания от ДНК, т. е. к нарушению структуры соответствующего гена.

У всех видов, размножающихся половым путем, потомству передаются лишь те мутации, которые затрагивают половые клетки. Поэтому весьма существенны те случайные нарушения, которые происходят при формировании половых клеток, в мейозе. Эти нарушения могут затрагивать не только отдельные гены, но и хромосомы в целом. Отдельные гаметы могут получить хромосому с искаженной генной структурой или вообще недополучить какую-то хромосому. Возможно также образование гамет с лишними хромосомами.

Тепловое движение молекул вещества — не единственная причина появления мутаций. Исследования выявили целый ряд *внешних* факторов, вызывающих мутации. Подобные факторы называют *мутагенными*. К ним относятся некоторые химические вещества и различного рода излучения — рентгеновские лучи, быстрые заряженные частицы, пучки нейтронов и т. д.

Польза и вред мутаций. С точки зрения эволюции мутации, безусловно, полезны. Более того, они необходимы. Огромное разнообразие генов у каждого вида, а также многообразие существующих на Земле видов — все это есть следствие многочисленных мутаций, которые происходили на протяжении многих миллионов лет (происходят и поныне). С точки зрения отдельных организмов мутации, как правило, вредны, в отдельных случаях даже

смертельны. Как следствие длительиой эволюции, оргаиизм появляется иа свет со сложным геиотипом, достаточио хорошо приспособлениым к условиям обитания. Случайиое измеиеиие геиотипа скорее всего вызовет какие-то иарушеиия в отлажеиином биологнческом механизме.

Мы вндим, таким образом, что мутации одиовремеиио и полезны (даже необходимы), и вредиы. Если у даииого вида мутации будут возникать слишком часто (иапример, в результате радиоактнвного заражеиия среды обитаиия), то это приведет к повышеиию смертностн оргаиизмов и, как следствие, к сокращеиию, а возможно, н к гнбели вида. Если у даииого вида мутации, иапротнв, проиcходят слишком редко, то при каком-иибудь зиачительном изменениn внешних условий даииый вид не сможет приспособиться н также погибиет. Например, мамонты ие сумели приспособиться к резкому похолодаиию во время ледиикового периода н вымерлн. Итак, плохо, когда мутаций очеиь миого, когда они пронсходят очень часто. Плохо также и когда мутаций практически нет нлн оии происходят слишком редко.

Организм и мутации. Приспособлеиие оргаиизма к условиям обитання предполагает также и приспособление к мутациям, вследствие чего степень вреда, приносимого мутацией, существеиио сиижается. Такое приспособлеине естествеиио, поскольку развитие вида иепосредственно связаио со степеиью выживаемости его представителей.

Обсудим этот вопрос с позиций геиетнки. Допустим, что даиная зигота возиикла в результате соедииеиия иормальиой и смутировавшей гамет. Говоря о смутировавшей гамете, будем полагать, что в какой-то хромосоме имеется испорченый (смутировавший) геи. Пусть этот ген отвечает за жизиенно важные для оргаиизма процессы, так что речь идет о действительио опасной мутации. Смутнровавшему гену противостоит нормальный геи в пáриой хромосоме. Смутнровавший геи может оказаться либо домииантиым, лнбо рецессивиым по отиошеиию к иормальиому геиу. Рассмотрим обе возможностн.

Еслн смутнровавшнй геи *доминантен*, то ои иемедлеиио начиет свою «вредную деятельиость», в результате которой оргаиизм погнбает уже в эмбрноиальиом периоде развития. Дарвниский отбор выполняет здесь свою саиитариую миссню задолго до того, как домннантная мутация распространнтся в потомстве. В результате не пронсходит иакоплеиия домииаитных смутировавших генов. Иное дело, если смутировавший геи *рецессивен*. Его подавляет нормальный ген, поэтому даниый оргаиизм оказывается феиотипическн здоровым. Более того, и в его потомстве будут появляться здоровые органнзмы-феиотипы. Лишь в нсключительио редких случаях рецессивный смутировавший геи может «заявить о себе» тогда, когда к какому-иибудь потомку этот ген попадет одиовременно н через отцовскую, и через материнскую гамету.

Так н хочется сказать, что мудрая природа позаботилась о том, чтобы умеиьшить опасиость губительных мутаций. Будем, одиако,

помнить, что природа ни о чем и ни о ком не заботится. Все дело в великом принципе отбора наиболее приспособленных. Иной «мудрости» у природы нет.

К сожалению, люди сами подчас способствуют повышению опасности мутации. Вероятность встречи в каком-то потомке двух рецессивных смутировавших генов увеличивается при браках между родственниками, а также браках, заключаемых в пределах какой-либо *ограниченной группы* людей, например в пределах одной общины, одной секты, затерянного где-нибудь в горах селения и т. д. Там, где практикуются подобные браки, неизбежно наблюдаются вспышки различных наследственных заболеваний (их называют *рецессивными заболеваниями*). Таких заболеваний известно около пятисот. Они могут вызывать идиотизм, слабоумие, глухонемоту, физическую неполноценность. Таким образом, всякое искусственное разъединение людей, разбиение их на замкнутые группы увеличивает генетическую опасность, приводя к повышению вероятности рецессивных заболеваний.

Во второй половине нашего века мутационная опасность резко возросла вследствие испытаний ядерного оружия. Радиоактивное излучение является сильно действующим мутагенным фактором. Поэтому трудно переоценить важность заключенного по инициативе Советского Союза международного «Договора о запрещении испытаний ядерного оружия в атмосфере, космосе и под водой». В 1963 году этот Договор подписали СССР, США, Великобритания. В настоящее время к нему присоединились уже более ста стран.

Закон гомологических рядов в наследственной изменчивости. Каждая отдельная мутация — явление случайное, ненаправленное, непредсказуемое. Если же у данного вида происходит относительно много мутаций (это лучше всего наблюдать у растений), то картина мутаций в целом обнаруживает *закономерность*, необходимость. Это подтверждает *закон гомологических рядов* в наследственной изменчивости, открытый видным советским биологом академиком Н. И. Вавиловым (1887—1943). На основании большого фактического материала Вавилов пришел к выводу, что генетически близкие виды должны характеризоваться сходными (гомологическими) рядами наследуемых измененных признаков. Если, например, мутации вызывают ряд каких-то довольно часто встречающихся наследственных признаков у ржи, то аналогичный ряд признаков должен наблюдаться также у пшеницы, ячменя, овса и некоторых других культур.

Открытый Вавиловым закон иногда сопоставляют с периодической системой элементов Менделеева, желая тем самым подчеркнуть, что, подобно системе Менделеева, этот закон позволяет предсказывать новые мутанты. В 1917 году во время научной экспедиции Вавилов нашел на Памире разновидность пшеницы с листьями, у основания которых не было лигулы (язычка). В то время биологи не знали ни безлигульной ржи, ни безлигульного ячменя. Но, по закону Вавилова, такие разновидности ржи и ячменя должны были встречаться. И вот в 1918 году была найдена

безлигульная рожь. Позднее, в 1935 году, был получен безлигульный ячмень в результате облучения рентгеновскими лучами обычного ячменя.

Эволюция глазами генетика

Было время, когда некоторые биологи пытались противопоставлять учения Дарвина и Менделя. Такое противопоставление следует отнести в разряд наиболее досадных заблуждений. Сегодня оно представляется абсурдным. Общепризнано, что именно генетика дала строгое научное обоснование дарвинской теории происхождения и эволюции видов, именно генетика разъяснила, как происходит наследование измененных признаков. Сегодня дарвинизм — это логически стройная, авторитетная наука, способная давать ценные практические рекомендации. Всеми своими корнями эта наука уходит в современную генетику.

Ненаправленная наследственная изменчивость. По выражению советского биолога академика И. И. Шмальгаузена, каждый вид и каждая его популяция таят в себе «резерв наследственной изменчивости». Этот резерв может быть использован через естественный отбор при изменении условий обитания.

Существуют два основных «механизма» появления ненаправленной наследственной изменчивости. Это, прежде всего, *мутационная изменчивость*. В конечном счете именно мутации лежат в основе наблюдаемого многообразия видов и многообразия генов внутри вида. Мутационные изменения совершаются очень медленно, но они совершаются непрестанно и с очень давних пор. Более оперативен «механизм» появления наследственной изменчивости в результате случайного *комбинирования родительских генов* при скрещивании. При этом надо различать комбинирование генов в результате того, что соединяются случайные пары разнополых гамет, и комбинирование генов в результате того, что в гамету попадают случайно перетасованные части парных хромосом (явление перекреста хромосом).

Разумеется, изменения при комбинировании генов ограничены рамками существующего фонда генов. Фонд этот, однако, огромен. Подсчитано, что из фонда генов отца и матери можно в принципе сконструировать до 10^{50} разных человеческих генотипов. Это невообразимо огромное число. На Земле живет менее 10^{10} человек. Можно утверждать, что два человека практически не имеют никаких шансов оказаться генетически тождественными (за исключением, конечно, близнецов, развившихся из одной зиготы). Каждый человек *генетически* уникален; он обладает единственным в мире генотипом.

«Демон Дарвина» против «Демона Максвелла». В четвертой главе мы познакомились с «демоном Максвелла». Не получая извне информации, этот «демон» принципиально не мог вершить отбор — не мог отобрать из одной половины сосуда более быстрые молекулы и пропустить их в другую половину. Беспомощность «демона

Максвелла» демонстрировала принципиальную *невозможность отбора на атомно-молекулярном уровне* — в полном соответствии со вторым началом термодинамики.

Говоря о происходящем в живой природе естественном отборе, американский биохимик и писатель-фантаст Айзек Азимов употребил термин «демон Дарвина». В отличие от беспомощного «демона Максвелла» этот «демон», напротив, действует весьма успешно, отбирая и пропуская в следующее поколение организмы с более высокими шансами на выживание и дальнейшее размножение. В чем же принципиальное отличие «демона Дарвина» от «демона Максвелла»? Ответ прост: они действуют на *разных* уровнях. Все начинается на атомно-молекулярном уровне. На этом уровне возникают случайные ненаправленные мутации, происходит случайная перекомбинация генов. Если бы «демон Максвелла» мог функционировать, то он тут же бы принялся отбирать наиболее «выгодные» мутации, наиболее «удачные» комбинации генов. Этого не происходит, поскольку на атомно-молекулярном уровне отбор невозможен.

И вот тут вступает в действие *принцип усиления*. Допустим, что в зиготу попал смутировавший ген. По мере развития организма происходят многократные деления клеток и в итоге ген-мутант оказывается продублированным примерно 10^{15} раз. Точно так же оказывается продублированной и реализовавшаяся в рассматриваемой зиготе случайная комбинация генов. Таким образом, в процессе становления фенотипа *случайные изменения* генетической программы оказываются *многократно усиленными*. Тем самым совершается переход с атомно-молекулярного уровня на уровень *макроявлений*. А на этом уровне отбор возможен. Подчеркнем: «демон Дарвина» не пытается заниматься отбором самих измененных генетических программ, он не уподобляется «демону Максвелла». Он действует на организмы-фенотипы, в которых любое изменение генетической программы оказывается увеличенным в миллионы миллиардов раз.

По-видимому, нет необходимости объяснять, как именно действует «демон Дарвина». Формы, в которых реализуется естественный отбор, описаны во всех учебниках биологии. Заметим лишь, что этот «демон» выглядит довольно неумолимым. Он действует жестко: уничтожает те фенотипы, которые случайно оказались неприспособленными, а из тех, которые оказались в той или иной мере приспособленными к условиям обитания, отдает предпочтение более приспособленным, а менее приспособленных, как правило, также уничтожает.

Впрочем, «демон Дарвина» действует не столь прямолинейно, предоставляя испытуемым лишний шанс. Не пригодившиеся сегодня изменения генетической программы могут пригодиться завтра. Сегодня они бесполезны и даже вредны, завтра они могут оказаться полезными. Значит, не надо торопиться с вынесением приговора. Пусть случайно возникшее изменение в генетической программе в течение нескольких поколений фенотипов «подремлет»,

замаскировавшись в рецессивном гене. Вдруг это пригодится в дальнейшем.

Разумеется, эффект «демона Дарвина», или, иными словами, естественный отбор, ни в коей мере не противоречит второму началу термодинамики. Как уже отмечалось, живые организмы существуют лишь благодаря притоку негэнтропии из окружающей среды, т. е. за счет повышения энтропии в этой среде. Этим повышением энтропии и приходится «расплачиваться» за действия «демона Дарвина».

Многообразие видов. Наблюдаемое на Земле многообразие видов, где наряду с простейшими сосуществуют и очень сложные, высокоорганизованные, есть результат эволюции, продолжающейся в течение вот уже более двух миллиардов лет. В тот неимоверно удаленный период на Земле обитали лишь некоторые виды бактерий и сине-зеленых водорослей. Через несколько сотен миллионов лет появились одноклеточные организмы с оформленным внутриклеточным ядром. Еще через несколько сотен миллионов лет возникли кишечнополостные, черви, моллюски. Примерно полмиллиарда лет назад появились рыбы, позднее земноводные и еще позднее рептилии. Около ста миллионов лет назад появились млекопитающие. Исследуя процесс эволюции, нетрудно обратить внимание на то, что здесь нет простого перехода от менее сложных видов к более сложным. Конечно, какие-то виды (и их было немало) отмирали; тем не менее сейчас можно видеть наряду со сложными видами и огромное количество простых. Эволюция шла не в направлении от *простого к сложному*, а в направлении *от менее приспособленного к более приспособленному,* поскольку именно в этом (и ни в каком другом) направлении действует естественный отбор. Характерная черта такого процесса — *увеличение числа видов,* все большее и большее их разнообразие. Естественно, что при этом будут появляться и все более организованные виды, придавая эволюционному процессу прогрессивный характер.

Можно указать ряд причин, объясняющих, почему эволюция приводит к увеличению числа различных видов. Во-первых, со временем возрастает наследственная изменчивость — накапливаются мутации, расширяется фонд генов. Во-вторых, при любом изменении условий имеется большое число вариантов приспособления. Естественный отбор пропускает любые приемлемые варианты. При этом могут быть отобраны варианты как с более сложной, так и с менее сложной организацией. В-третьих, возникнув однажды, вид обнаруживает устойчивость. В частности, он противостоит опасности растворения в других видах. Напомним, что при скрещивании между разными видами гибриды не могут образовать половые клетки, а следовательно, не могут иметь потомство. Разумеется, рассматривая процесс увеличения числа видов, надо учитывать и обратные процессы, например уничтожение вида в результате межвидовой борьбы или гибель вида из-за неспособности приспособиться к внезапно и очень резко изменившимся условиям обитания.

Непредсказуемость новых видов. В четвертой главе мы рассматривали флуктуации в коллективе молекул газа и убедились, что флуктуации величин, относящихся к отдельной молекуле, велики. Они сопоставимы со средними значениями величин. Флуктуации же величин, характеризующих макросистему, напротив, крайне малы. Поэтому макросистему можно описывать на основе не вероятностных, а динамических законов (что и делается в термодинамике). Получается, что при переходе с атомно-молекулярного уровня рассмотрения на макроуровень происходит как бы взаимная компенсация многочисленных случайных отклонений в поведении отдельных молекул. В результате поведение макросистемы как целого становится однозначно предсказуемым.

В живой природе мы встречаемся с качественно иной ситуацией. Отдельные флуктуации, характеризующие случайные изменения той или иной генетической программы, усиливаются в миллионы миллиардов раз и обнаруживаются на макроуровне — в организме-фенотипе. Никакой взаимной компенсации подобных флуктуаций здесь нет. *Каждая флуктуация вырастает до макроразмеров*. Поэтому можно утверждать, что процесс эволюции в живой природе является *принципиально непредсказуемым* в том смысле, что нельзя предвидеть возникновение того или иного конкретного вида. Иначе говоря, любой вид оказывается явлением случайного характера. Его можно уничтожить, можно создать какой-нибудь новый вид, но нельзя восстановить исчезнувший вид. В этом смысле любой из существующих ныне видов уникален.

Заключение. Мы обсудили ряд вопросов биологии, связанных с генетикой и эволюционной теорией. Именно в этих вопросах особенно ярко проявляется *фундаментальность вероятностных закономерностей*, именно здесь наиболее отчетливо проступает *принципиальная роль случайностей*. И все же тема «Вероятность в биологии» много шире. Она включает в себя также ряд проблем, которым не нашлось места в данной книге. К подобным проблемам относятся, например, проблема возникновения жизни на Земле, проблема изменения численности популяций, проблема моделирования процессов в нервной системе, проблема создания модели человеческого мозга и ряд других.

Заключительная беседа

Только кончая задуманное сочинение, мы уясняем себе, с чего нам следовало его начать.

Блез Паскаль

Этот удивительно симметричный мир, построенный на вероятности

Беседуя с читателем, автор полагает, что тому знакома ранее изданная книга автора «Этот удивительно симметричный мир» (М.: Просвещение, 1982), где была предпринята попытка проанализировать понятие *симметрии* и показать, что представления о симметрии и асимметрии лежат в основе физической картины мира.

Автор. Книга о мире вероятностей закончена. Надеюсь, что она дала вам немало пищи для размышлений.

Читатель. Должен признаться, что некоторые моменты все же не укладываются в сознание. Например, мне трудно принять идею использования случайного для решения тех или иных проблем. Я имею в виду метод Монте-Карло, принцип действия гомеостата, персептрон. Все это похоже на некоторое «чудо».

Автор. А между тем все это не большее «чудо», чем таблица случайных чисел.

Читатель. Я не вполне понимаю вас.

Автор. Каждая очередная цифра в таблице появляется независимо от того, какие цифры появились раньше. И несмотря на это таблица в целом обнаруживает устойчивость. Цифры выпадают независимо друг от друга, а в то же время частота появления любой цифры оказывается вполне определенной.

Кстати, бесполезно пытаться написать набор случайных цифр, что называется от руки. Вот вы начинаете писать, например, 8, 2, 3, 2, 4, 5, 8, 7... И, конечно, ло́вите себя на мысли, что надо бы написать 1 и 6, так как они еще не появлялись. Вы непроизвольно корректируете свои последующие действия в зависимости от предыдущих. В итоге таблица случайных чисел у вас не получится.

Важно сознавать, что появление очередного случайного события никоим образом не связано с появлением предыдущих событий. Поэтому и кажется «чудом» устойчивость, наблюдаемая в картине из большого числа случайных событий. В конечном счете именно из этого «чуда» вытекают чудесные свойства персептрона или метода Монте-Карло.

Читатель. Я могу согласиться, что «корень зла» таится в конечном счете в таблице случайных чисел. Чем же объясняются загадочные свойства этой таблицы?

Автор. Объяснение заключается в слове «симметрия».
Читатель. Поясните, пожалуйста.
Автор. Выявляя при составлении таблицы очередную цифру, вы заботитесь о том, чтобы была обеспечена *симметрия* по отношению к выпадению любой цифры. Иначе говоря, любая цифра от нуля до девятки должна иметь *одинаковые шансы* появиться.
Читатель. Предположим, что я вытаскиваю из мешка шары, помеченные разными цифрами. Какую симметрию вы здесь имеете в виду?
Автор. Например, симметрию по отношению к взаимному обмену шаров. Вообразите, что все шары вдруг поменялись местами. Если указанная симметрия есть, то вы не заметите совершившегося обмена. Но это еще не все. Возвращая всякий раз шары обратно в мешок и тщательно перемешивая их, вы тем самым как бы восстанавливаете исходную ситуацию и заботитесь о том, чтобы вся система оказалась симметричной по отношению к переходу от одного акта вынимания шара к другому такому акту. Как видите, объяснение оказывается достаточно серьезным. Симметрия и асимметрия относятся к наиболее фундаментальным понятиям. Эти понятия лежат в самой основе естественнонаучной картины мира.
Читатель. Я читал вашу книгу «Этот удивительно симметричный мир». Меня удивило, насколько глубоко проникает симметрия во все явления, происходящие в нашем мире. Теперь я вижу, что то же самое можно сказать о случайности.
Автор. Фактически в той книге речь шла не просто о симметрии, а о *диалектическом единстве симметрии и асимметрии.* Соответственно и здесь мы рассматривали не просто случайность, а *диалектическое единство необходимого и случайного,* которое, кстати, и выражается через вероятность.
Читатель. Судя по сделанным ранее замечаниям, между необходимостью-случайностью и симметрией-асимметрией существует связь.
Автор. И весьма глубокая. Принципы симметрии-асимметрии управляют законами природы, как и законами человеческого творчества. Не менее принципиальна и ва́жна роль вероятностных принципов.
Читатель. Хотелось бы более конкретно обсудить связь между симметрией и вероятностью.
Автор. Как известно, классическое определение вероятности основывается на подсчете равновозможных исходов. В свою очередь, *равновозможные исходы всегда связаны с определенной симметрией.* С равновозможными исходами мы встречались не только тогда, когда занимались подбрасыванием кубика или монеты. Вспомните определение статистического веса макросостояния через число равновозможных микросостояний (глава четвертая). Вспомните обсуждение равновозможных альтернатив при рассмотрении законов Менделя (глава шестая). Во всех этих случаях вероятность некоторого события определялась как величина, пропор-

циональная числу равновозможных (можно сказать, симметричных) исходов, в каждом из которых реализуется данное событие. Иными словами, вероятность события есть сумма вероятностей соответствующих равновозможных исходов.

Читатель. Я начинаю думать, что и само правило *сложения вероятностей* основано на некоей симметрии.

Автор. Интересная мысль.

Читатель. Мы ищем вероятность того, что произойдет либо первое событие, либо второе, причем безразлично, какое именно, поскольку любое из них приводит к результату. Симметрия связана здесь с независимостью получения результата по отношению к замене одного события другим.

Автор. Можно пойти дальше. Предположим, что имеется еще более глубокая симметрия, связанная с *неразличимостью* первого и второго события (подобные ситуации мы обсуждали в пятой главе). В этом случае правило *сложения вероятностей* заменяется правилом *сложения амплитуд вероятностей.*

Читатель. Действительно, связь между симметрией и вероятностью явно просматривается.

Автор. Эту связь можно представить в еще более конкретном виде, если воспользоваться понятием *информации.* Вы помните, конечно, что информация принципиально основывается на вероятности (см. главу третью). Связь же информации с симметрией такова: *более симметричному состоянию соответствует меньшая информация.*

Читатель. Тогда можно утверждать, что с повышением симметрии состояния должна возрастать его энтропия?

Автор. Именно так. Взгляните на рисунок 4.12. Состояние с наибольшим статистическим весом и, следовательно, с наибольшей энтропией — это состояние, соответствующее равномерному распределению молекул по обеим половинам сосуда. Оно, очевидно, и наиболее симметрично (симметрия по отношению к отражению в мысленном зеркале, плоскость которого делит сосуд на две половины).

Читатель. Тут есть над чем подумать. Получается, что в процессе творчества человек понижает симметрию. Однако симметрия сама по себе широко используется в творчестве. Нет ли здесь противоречия?

Автор. Противоречия здесь нет. Ведь человек использует в своем творчестве не просто симметрию, а симметрию-асимметрию. Об этом мы уже говорили. Конечно, поднимаемые вопросы требуют специального рассмотрения. Здесь же мы можем слегка коснуться некоторых проблем, не входя в подробности.

В книге о симметрии мы подчеркивали, что симметрия действует в направлении *ограничения числа возможных вариантов структур, вариантов поведения.* Очевидно, что необходимость действует в *том же самом* направлении. С другой стороны, асимметрия действует в направлении увеличения числа возможных вариантов. В этом же направлении действует и случайность. Выше мы не-

однократно обращали внимание на то, что случайности *создают новые возможности*, порождают новые альтернативы.

Читатель. Значит, можно говорить о следующей «расстановке сил». На одной стороне симметрия и необходимость. На другой — асимметрия и случайность.

Автор. Да, именно такова «расстановка сил». В заключение мне хотелось бы вспомнить притчу о буридановом осле. Именно с этой притчи, если вы не забыли, мы и начали нашу первую беседу в книге «Этот удивительно симметричный мир».

Читатель. Я хорошо помню эту притчу. Некий философ, которого звали Буридан, уезжая, оставил своему ослу две одинаковые охапки сена. Осел не смог решить, с какой охапки начинать, и умер с голоду.

Автор. Притча рассматривалась как пример зеркальной симметрии. Представьте себе картину: две одинаковые охапки сена и посредине между ними осел, который не в состоянии предпочесть одну охапку другой.

Читатель. Как я понимаю, осла погубила симметрия.

Автор. Согласно притче, это так. В действительности же осел живет не просто в «симметричном мире», а в «симметричном мире, построенном на вероятности». Какая-либо незначительная случайность (на осла села муха, осел вздрогнул или просто чуть-чуть пошевелился) легко *разрушает симметрию* — одна из охапок сена оказывается немного ближе, чем другая. Вместе с тем исчезает и проблема выбора. Здесь произошло, как говорят физики, *спонтанное нарушение симметрии.*

Читатель. Можно ли отсюда заключить, что симметрия губительна, а случай спасителен?

Автор. Уверен, что вы сами понимаете излишнюю категоричность такого вопроса. В свое время мы убедились, что симметрия уменьшает число вариантов поведения, сокращает альтернативы. Логично допустить, что это уменьшение может привести к безвыходной ситуации, может завести в тупик. И тогда жизненно важна спасительная случайность. С другой стороны, чрезмерность случайностей, обилие разнообразных вариантов, существенная разупорядоченность — все это также может оказаться губительным. И тогда на помощь приходит упорядочивание, т. е. на помощь приходят симметрия и необходимость.

Читатель. Опасность со стороны случайностей понятна. Однако какая может быть опасность со стороны симметрии? Если, конечно, мы не уподобимся буриданову ослу.

Автор. Во-первых, не отмахивайтесь от осла. Пример с ослом был выбран не в качестве иллюстрации из жизни животных, а для того, чтобы продемонстрировать некоторую проблему. Во-вторых, совсем нетрудно привести практический пример *опасности симметрии.* Строители современных мостов, высотных зданий, башен знают, что конструкция не должна быть безупречно симметричной из-за опасности возникновения резонансных колебаний, которые могут привести к разрушению конструкции. Известны случаи раз-

рушения мостов вследствие резонанса, вызванного, например, ротой маршировавших по мосту солдат, ритмичными порывами ветра или другими внешне безобидными причинами. Поэтому при строительстве больших сооружений всегда незначительно нарушают симметрию конструкции за счет того, что случайным образом вводят в нее отдельные асимметричные балки, консоли, плиты и т. п.

Читатель. Действительно, симметрия может оказаться опасной. Насколько я понял, совсем нетрудно нарушить симметрию. Достаточно случайной мухи, которая села на осла, или лишней балки в конструкции.

Автор. Вы обратили внимание на важное обстоятельство. Именно *неустойчивость симметрии* позволяет легко нарушить ее и, в частности, создает возможность ее спонтанного нарушения.

Читатель. Неустойчивая симметрия. Это нечто новое.

Автор. Исследования неустойчивости симметрии имеют очень короткую историю — всего десять лет. Они привели к возникновению нового научного направления, называемого *теорией катастроф*. Эта теория изучает взаимосвязи симметрии и случайности с точки зрения развития различных процессов и явлений.

Читатель. Название теории звучит несколько мрачновато.

Автор. Рассматриваемые здесь катастрофы совершаются на самых разных уровнях. Предположим, что частица вызывает бурный процесс в счетчике Гейгера — Мюллера, в результате чего она и регистрируется. Этот процесс есть катастрофа в масштабах микромира. Огромный мост или современный реактивный самолет внезапно разваливаются вследствие возникших в их конструкциях резонансных колебаний. Это есть пример катастрофы уже в привычных для нас масштабах. Примеры катастроф могут быть достаточно разнообразными — внезапная кристаллизация переохлажденной жидкости, рождение горного обвала, возникновение генерации излучения в лазере. Во всех подобных случаях система характеризуется неустойчивой симметрией, которая может разрушиться под действием различного рода случайных факторов. Эти случайные факторы могут оказывать весьма незначительное воздействие, могут являться, казалось бы, совершенно безобидными. Но они разрушают симметрию и тем самым *развязывают в неустойчивой системе бурно протекающие процессы*, которые могут рассматриваться как своего рода катастрофы.

Читатель. По-видимому, именно в теории катастроф особенно четко проявляется вся глубина связи между симметрией-асимметрией и необходимостью-случайностью.

Автор. Вполне с вами согласен. Впрочем, все это — тема уже иной книги.

Литература

К главе 1

Реньи А. Письма о вероятности: Пер. с венг.— М.: Мир, 1970.
Глеман М., Варга Т. Вероятность в играх и развлечениях (элементы теории вероятностей в курсе средней школы): Пер. с франц.— М.: Просвещение, 1979.
Чубарев А. М., Холодный В. С. Невероятная вероятность.— М.: Знание, 1976.
Хургин Я. И. Как объять необъятное.— М.: Знание, 1979.
Вентцель Е. С. Теория вероятностей (первые шаги).— М.: Знание, 1977.

К главе 2

Вентцель Е. С. Исследование операций — задачи, принципы, методология.— М.: Наука, 1980.
Вентцель Е. С. Элементы теории игр.— М.: Физматгиз, 1961.
Растригин Л. А. Этот случайный, случайный, случайный мир.— М.: Молодая гвардия, 1974.
Платонов Г. А., Файнберг М. А., Штильман М. С. Поезда, пассажиры и... математика.— М.: Транспорт, 1977.
Хургин Я. И. Да, нет или может быть...— М.: Наука, 1977.

К главе 3

Растригин Л. А. Этот случайный, случайный, случайный мир.— М.: Молодая гвардия, 1974.
Пекелис В. Д. Маленькая энциклопедия о большой кибернетике.— М.: Детская литература, 1970.
Теплов Л. П. Очерки о кибернетике.— М.: Московский рабочий, 1963.
Реньи А. Трилогия о математике: Пер. с венг.— М.: Мир, 1980.
Артоболевский И. И., Кобринский А. Е. Знакомьтесь — роботы! — М.: Молодая гвардия, 1979.

К главе 4

Смородинский Я. А. Температура.— М.: Наука, 1981.
Компанеец А. С. Законы физической статистики.— М.: Наука, 1976.
Шамбадаль П. Развитие и приложение понятия энтропии: Пер. с франц.— М.: Наука, 1967.

К главе 5

Компанеец А. С. Что такое квантовая механика? — М.: Наука, 1977.
Пономарев Л. И. По ту сторону кванта.— М.: Молодая гвардия, 1971.
Каройхази Ф. Истинное волшебство: Пер. с венг.— М.: Атомиздат, 1980.
Батыгин В. В. Законы микромира.— М.: Просвещение, 1981.

К главе 6

Полынин В. М. Мама, папа и я.— М.: Советская Россия, 1967.
Лучник Н. В. Почему я похож на папу.— М.: Молодая гвардия, 1966.
Шредингер Э. Что такое жизнь?: Пер. с англ.— М.: Атомиздат, 1972.
Резник С. Е. Николай Вавилов.— М.: Молодая гвардия, 1968.
Медников Б. М. Аксиомы биологии.— М.: Знание, 1982.

Оглавление

Лев Васильевич Тарасов

Мир, построенный на вероятности

Редактор
Н. В. Хрусталь

Макет и оформление
В. А. Крючкова

Художники
В. А. Крючков, О. М. Шмелев

Художественный редактор
В. М. Прокофьев

Технический редактор
В. Ф. Коскина

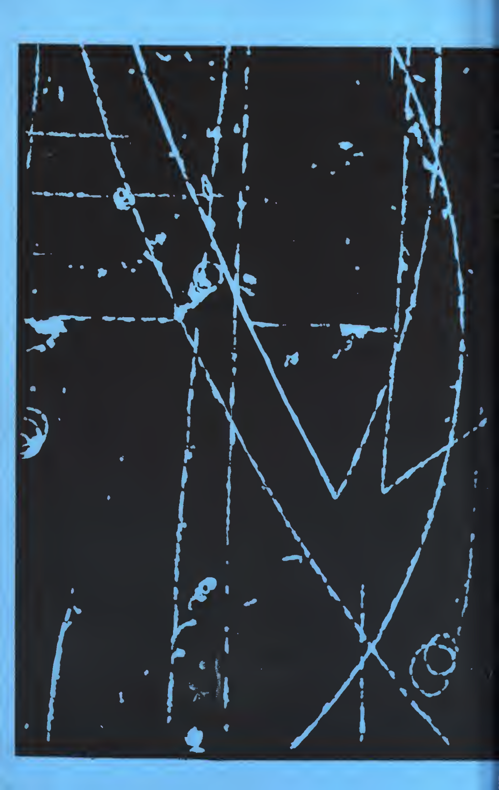
Корректор *Г. И. Вольсон*

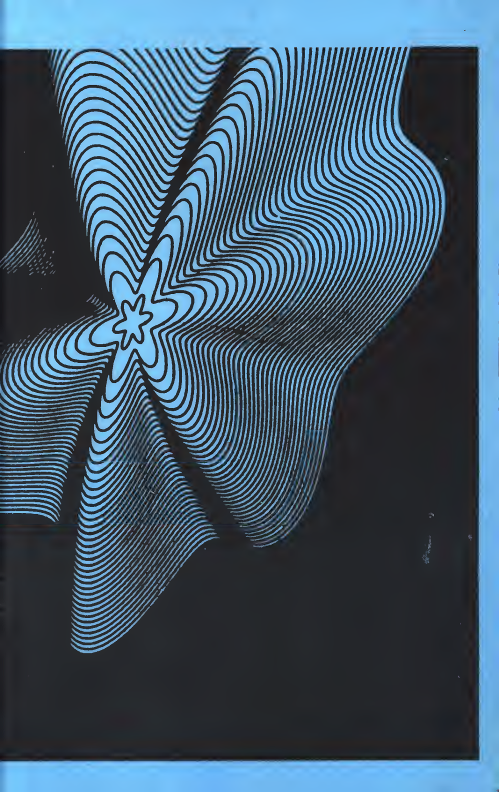
ИБ № 6927

Сдано в набор 21 06 83. Подписано к печати 10.07 84 Формат $60\times90^{1}/_{16}$ Бум офсетная № 1. Гарнит. литературная. Печать офсетная. Усл печ л 12+форзац 0,25 Усл. кр-отт. 45,5. Уч-изд л. 13,02+форзац 0,33 Тираж 230 000 экз. Заказ № 664. Цена 1 руб. 10 коп.

Ордена Трудового Красного Знамени издательство «Просвещение» Государственного комитета РСФСР по делам издательств, полиграфии и книжной торговли 129846, Москва, 3-й проезд Марьиной рощи, 41.

Смоленский полиграфкомбинат Росглавполиграфпрома Государственного комитета РСФСР по делам издательств, полиграфии и книжной торговли. Смоленск 20. ул Смольянинова, 1





1 р. 70 к.